УДК 81’33 DOI: 10.18698/2541-8009-2022-04-784

РОЛЬ КОНТЕКСТА ПРИ АВТОМАТИЧЕСКОМ РАСПОЗНАВАНИИ УСТНОЙ РЕЧИ

Н.В. Кокурина kokurina.natasha@gmail.com
Д.М. Жуков dmzhukov@outlook.com

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Рассмотрен контекст как фактор влияния на результаты автоматического распознавания речи. Обоснована необходимость классификации контекстов при автоматической обработке естественного языка. Описаны структурный и неструктурный типы контекстов, выделен интонационный контекст. Рассмотрены лексический и синтаксический контексты, при которых значение лексической единицы зависит от ближайшего окружения. Каждый из выделенных контекстов, влияющих на качество автоматического распознавания устной речи, проиллюстрирован примерами с комментариями. Подчеркнуто, что особую значимость для автоматического распознавания речи играют семантический и ситуационный контексты. Выделен культурный контекст как наиболее сложный для автоматической обработки речи.

Ключевые слова

Контекст, распознавание речи, понимание речи, микроконтекст, макроконтекст, устная речь, лексический контекст, структурный контекст, синтаксический контекст, культурный контекст, акустика

Поступила в редакцию 22.10.2021

Введение. Автоматическое распознавание звуков речи — актуальная научная и техническая проблема, решение которой позволит значительно ослабить влияние негативных факторов при вводе в ЭВМ звуковой текстовой информации и организовать управление с помощью звуков речи. Проблема ввода устной речи рассматривается давно и сегодня решена на достаточно высоком уровне, однако имеет три принципиальных ограничения:

– персональное, поскольку автомат распознает язык только конкретного говорящего;

– связанное с условиями подготовки — автомат распознает язык только тогда, когда он заранее подготовлен;

– языковое — автомат распознает только ограниченное количество слов [1].

Отметим, что программные средства с такими ограничениями создают благоприятные условия для работы в автоматизированном управлении и информационном поиске, однако они непригодны для организации диалога с системой в массовом обслуживании на естественном языке. Для устранения этих ограничений нужно, чтобы автомат распознавал слова, отдельные звуки, реализованные любым говорящим. В настоящее время одной из главных технических

проблем является организация эффективного восприятия компьютером так называемой сплошной речи. В сплошной речи сложно определить, где заканчивается одно слово и начинается другое. При этом акустические образы произнесенных слов намного больше зависят от контекста. В системах распознавания изолированных слов этих проблем нет, поскольку слова разделены паузами [2].

Артикуляторный аппарат человека производит звуки речи. Каждый звук имеет свои акустические характеристики, которые можно описать набором физических параметров. Осмысленность звуков в речевом потоке реализуется в форме мелодического непрерывного речевого фрагмента, в котором заложены необходимые семантические признаки [3].

При попытке научить компьютер понимать смысл «услышанных» им слов сразу возникает ряд научных проблем и задач. Прежде всего нужно преобразовать колебания воздуха в электрические сигналы с помощью микрофона, отфильтровав при этом помехи и шумы. Далее каким-то образом сигнал необходимо представить в цифровой форме, доступной для обработки с помощью компьютера (оцифровать). Здесь есть разные возможности: можно вводить в компьютер информацию об амплитуде звукового сигнала, а можно анализировать спектральный состав сигнала, выделяя из сигнала набор основных частот. Эту информацию можно комбинировать [4].

Для выделения из оцифрованного звука лингвистических конструкций применяют различные математические методы в сочетании со специальным компьютерным оборудованием, таким, например, как аппаратные или программные нейронные сети [5]. На протяжении всей истории систем распознавания речи эти методы постоянно изменялись [1]. При этом одни методы устаревали по причине их неэффективности, а другие разрабатывались и совершенствовались [4].

Одно дело — научить компьютер распознавать отдельные фонемы и слова, и совсем другое — научить компьютер понимать смысл сказанного. Без понимания контекста произносимых слов их правильная интерпретация не всегда возможна [4]. Исследования в области лингвистики и семантики речи позволили прийти к выводу, что восприятие речевых образов происходит на нескольких уровнях. Неопределенность или ошибки, возникающие на более низком уровне (по причине недостаточности акустической информации), обнаруживаются и исправляются на высших уровнях благодаря имеющимся у слушателя знаниям языка и предмета разговора, т. е. путем привлечения лингвистической и семантической информации [1].

Цель статьи — исследовать контекст как фактор влияния на результаты автоматического распознавания речи.

Роль контекста при автоматическом распознавании устной речи. Человек, воспринимая сказанное, стремится узнать речевые отрезки как можно большей

длины. Такой степени принято считать фонетическое слово. Фонемный образ слова в ходе восприятия может быть определен двумя путями:

1) слушатель может, пользуясь признаками слова как минимальной единицей восприятия, отнести его к определенному классу понятий. Этого достаточно в условиях стесненной ситуации, установки;

2) в случае широкого поля значение слова, когда оно не может быть причислено к какому-либо классу, слушатель должен перейти к детальному анализу звукового оформления слова путем выявления ключевых признаков [1].

Наиболее вероятным является понимание восприятия как адаптивного (приспосабливаемого) процесса, в котором процедура восприятия информации подчинена средствам выделения сигнала и цели слушания.

Если слушатель способен обнаружить лингвистические закономерности в звуках, его слуховой аппарат может использовать информацию, что накапливается во время речевого акта, для принятия дискретных решений относительно определения тех или иных завершенных звуковых конструкций (слов, фраз). Иначе процесс принятия решения в большей степени опирается на текущие акустические параметры, т. е. информация о сигнал добывается путем непрерывного сравнения и выбора приемлемого акустического эталона. Отсюда следует, что размеры элементов восприятия (единицы языка) меняются в зависимости от целей распознавания, а скорость обработки сигналов варьируется в соответствии с типом поступающей информации.

В наше время принята теория речевого восприятия последовательностей, согласно которой восприятие подразделяется на два этапа: автономный и неавтономный, т. е. этап взаимодействия информации различных уровней.

На первом этапе восприятия при поступлении акустико-фонетической информации о слове активизируются все слова, начинающиеся с одной и той же фонемы. Например, при произношении слова *сын* должны активизироваться и другие слова с фонемой [*с*], такие как *соло*, *сода*, *смех*, *сыр* и т. д. Слова, активизированные на основе первой фонемы из цепочки фонем, составляющих данное слово, образуют последовательность или группу. Этот процесс выбора слов происходит автономно, потому что используется только акустико-фонетическая информация о слове. Далее следует процесс активизации сопоставления слов-кандидатов на основе услышанного на уровне акустико-фонетических параметров. При этом включаются все источники речевой информации для того, чтобы отобрать нужное слово. На основе полученной информации слова типа *соло*, *сода*, *смех* могут быть удалены из последовательности, останутся только слова на [*сы*], такие, как *сын, сыр*. Однако слово *сыр* несовместимо с предварительной семантической и синтаксической информацией, оно будет исключено из данной последовательности. Далее в процесс восприятия устной речи после распознавания слова вступает контекст.

Установлено, что слушатель распознает слово в изолированном виде или в контексте после того, как прослушает его до конца. Момент узнавания слова зависит от целого ряда факторов, в том числе от физических характеристик слова, детально описанных в [3] (его продолжительности, качества звуковых стимулов), а также от лингвистических характеристик (частоты применяемости, количества слогов, количества других слов в лексиконе, похожих по звучанию на слово, которое человек пытается узнать), представленных в [6].

Под моментом распознавания слова понимают точку распознавания, начиная с которой эта последовательность сегментов ассоциируется с определенным словом. На этом этапе важную функцию выполняет контекст. Процесс восприятия зависит от двух факторов: фонетической информации, т. е. звукового оформления и предшествующего контекста, и информации речевых уровней. В связи с этим важно, каковы типы контекста, этап восприятия речи и каким образом осуществляется влияние на процесс коммуникации.

Под контекстом принято понимать семантико-грамматическое и коммуникативное единство определенного текстового элемента с текстовым и ситуативным окружением как индикатором значения и функционального веса этого элемента [2]. Различают микро- и макроконтекст. Микроконтекст — ближайшее окружение текста. Так, предложение получает смысл в контексте абзаца, абзац — в контексте главы и т. д. Макроконтекст — это вся система знаний, связанная с предметной областью, т. е. знания об особенностях и свойствах, явно не указанных в тексте. Другими словами, любое знание обретает смысл в контексте некоторого метазнания [2].

Типы контекстов при распознавании устной речи. При решении проблемы автоматического распознавания устной речи различают структурный и неструктурный типы контекстов. Структурный контекст может быть определен результатом применения ограничений, налагаемых на способы объединения элементов языка в единицы более высокого уровня [7]. Ограничения этого типа могут быть применены на уровне фонем, морфем, словосочетаний, высказываний и даже на уровне текста. Структурный контекст определяет, какие элементы могут быть объединены в структурные единицы, например:

Иди ко мне — и дико мне.

Пока лечилась — покалечилась.

Мы же на ты — мы женаты.

Ты жеребенок — ты же ребенок.

Несуразные вещи — несу разные вещи.

Ему же надо будет — ему жена добудет.

Надо ждать — надо ж дать.

Неструктурные типы контекста не влияют на отдельные элементы высказываний, а лишь указывают на определенные ассоциативные связи между слова-

ми. Например, *студент — университет.* Это слова близки по семантическим, а не по структурным признакам, например: *охрана* и *защита* — это синонимы, а *правоохранительные органы* и *правозащитные органы* — не синонимы; *бесчеловечно* и *безлюдно* — не синонимы.

Кроме вышеупомянутых типов контекста выделяют также контекст интонационный, лексический, синтаксический, семантический, ситуативный.

Громкость и темп речи постоянно изменяются. Более того, одна и та же фраза, сказанная разными людьми или даже одним человеком, находящимся в разных психических состояниях, может иметь разную спектрально-временную окраску. Это сильно затрудняет создание универсальных систем распознавания, «понимающих» речь разных людей [4]. Низкоуровневые изменения интонации и ритма могут полностью изменить значение слова. Примером является сарказм. Слово *уважаемый* можно сказать так, что это будет звучать как оскорбление.

Лексический контекст — это совокупность лексических единиц, слов и устойчивых словосочетаний, в окружении которых используется данная единица. Влияние лексического контекста заключается в том, что единицы лексического уровня влияют на распознавание единиц фонологического уровня. Например, фонема [*b*] в контексте [*iss*] производит слово, что имеет значение, тогда как фонема [*g*] в том же контексте продуцирует квазислово, т. е. слово, не наполненное смыслом.

Синтаксический контекст — это та синтаксическая конструкция, в которой употребляется данное слово, словосочетание или придаточное предложение.

Синтаксический контекст, как показывают теоретические и экспериментальные данные, в незначительной степени влияет на процесс восприятия речи, особенно на ранних этапах. В основном принадлежность слова к той или иной части речи можно определить с помощью суффиксов. Этот факт ограничивает влияние синтаксического контекста, но используется при распознавании суффиксов [5].

Контекст, основанный на информации о значении слова, может быть двух видов: семантический и ситуативный.

Под семантическим контекстом понимают любой тип контекста, который является смысловым. Этот смысл может быть основан, в том числе, на семантических ассоциативных связях или прагматическом значении. Слова, которые отвечают семантическому контексту, распознаются быстрее. Восприятие речи человеком в виде нормального линейного процесса — явление нераспространенное, поэтому человек в процессе восприятия пользуется методом эвристик, т. е. отражении действительности, что опережает семантический вид и позволяет сразу переходить к построению многоуровневых процедур с возможностью обратиться к высшим уровням языка. Человеку присущ активный поиск в пространстве текста с целью установления его смысловой структуры на максималь-

но ранних этапах процесса восприятия речи. Один из первых шагов — поиск темы сообщения.

В качестве примера рассмотрим слово *класс*, значение которого определяется в рамках предметной области. В естественных науках *класс* — произвольная совокупность множеств, обладающих каким-либо определенным свойством или признаком; в теории классификации — группа предметов или явлений, обладающих общими признаками: в программировании — абстрактный тип данных в объектно-ориентированном программировании, задающий общее поведение для группы объектов; модель объекта; в образовании — помещение для учебных занятий. При этом даже в рамках одной предметной области слово может иметь несколько значений, например, слово *класс* в культуре может иметь следующие значения:

«Класс» — советская/российская поп-группа, возглавляемая Олегом Кацурой;

«Класс» — эстонский фильм;

«Класс!» — российская телекомпания, основанная в 1994 г.;

класс (персонажа) — архетип персонажа в ролевых играх.

Отметим, что определение значения не представляет сложности для человека; в приведенных ниже примерах слова *кинофестиваль* и *телекомпания* указывают на точное значение слова *класс*.

В июне «Класс» был представлен на 42-м международном кинофестивале в Карловых Варах, получив два приза.

Среди полюбившихся многим телезрителям проектов телекомпании «Класс!» можно вспомнить программу для всей семьи «Без репетиций!», выходившую в эфир в 2004 году на канале ТВЦ.

Ситуативный (экстралингвистический) контекст включает обстановку, время и место, к которому относится высказывание, а также любые факты реальной действительности, знание которых помогает рецептору правильно интерпретировать значения языковых единиц в высказывании, например:

Разговор двух иностранцев:

— Я столько лет прожил в России, но так и не смог нормально выучить русский язык…

— А что случилось?

— Да подхожу я на рынке к продавцу и спрашиваю: «Это что?» Он мне отвечает: «Черная смородина». Спрашиваю: «А почему красная?» Он отвечает: «Да зеленая еще…»

Использование естественного языка в общении всегда диктуется целью говорящего. Эта цель сама по себе не предполагает точной формулировки, но указывает на скрытые ориентиры, которые помогают ограничить неоднозначность высказываний в меру необходимости решения соответствующей задачи. Способность использовать глобальный контекст для требуемого уменьшения неодно-

значности, не прибегая к явной формализации, связана с периферийным сознанием. Это краевое сознание учитывает скрытые ориентиры, заключенные в контексте, а также некоторые грамматические конструкции, и в конечном счете все это должно быть учтено и четко сформулировано для машины [8].

Произнесение и понимание предложения естественного языка предполагают достаточное знание того, как предложение связано с контекстом. Поэтому для того, чтобы заставить машину понимать и решать задачи речевого взаимодействия, необходимо научить машину связывать слова и словосочетания с ситуациями реальной действительности.

Рассмотрим фразу: «*Первым Николай распечатал письмо от Сони*». Как в этом случае понимать слово «распечатал»? Если знать, что это отрывок из романа «Война и мир», это слово должно означать «*открыл*», «*снял печать*», поскольку действие происходит в XIX в. Тогда не было принтеров, а письма передавались в конвертах, а чтобы их открыть, нужно было снять печать. Однако это же слово в этой же фразе в современном тексте вполне могло бы иметь смысл «*послать на печать*». Таким образом, интерпретатор, знающий происхождение текста и обладающий некоторыми знаниями истории технологий, поймет этот текст иначе, нежели тот, кто всеми этими знаниями не обладает [9]. При этом машинный переводчик в приведенной выше фразе слово «*распечатал*» переводит правильно, но только потому, что в сочетании со словом «*письмо*» слово «*распечатать*» он всегда переводит как *to open*. Поэтому фразу «*Он распечатал письмо, и принтер сломался*» машинный переводчик переводит некорректно [9].

Если речь идет о художественном тексте, то чаще всего автор рассчитывает на некий определенный набор знаний читателя — культурный контекст, — однако современные постмодернистские тексты вполне могут быть рассчитаны на различное, но эквивалентно правильное понимание носителями различных культурных контекстов.

Для человека обращаться к культурному контексту (или к базе знаний) настолько естественно, что он чаще всего этого не замечает. Однако иногда удивительно, сколь сложные логические построения требуются, чтобы доказать то, что любому человеку кажется очевидным, и компьютеру придется иметь дело именно с этими логическими построениями; от объема и сложности этих построений будет зависеть скорость понимания текста компьютером. Конечно, можно придумать облегченный язык, для понимания которого будет требоваться не более чем знание слов, использованных в тексте. Однако такой язык будет лишен многих свойств естественного языка [9].

Понимание языка — это не просто передача слов. Оно требует знаний о целях говорящего, контексте, а также о предметной области. При создании программ понимания естественного языка необходимо учитывать такие аспекты,

как немонотонность, изменение убеждений, иносказательность, возможность обучения, планирования и практическая сложность человеческих взаимоотношений [10].

Заключение. На сегодняшний день проблема ввода устной решена на достаточно высоком техническом уровне, однако имеется ряд принципиальных ограничений, препятствующих массовому обслуживанию: зависимость от диктора, ограниченный словарь. Восприятие речи системой обусловлено факторами контекста, при котором семантика и ситуация высказывания играют ключевую роль в адекватном распознавании. Также при создании программ понимания естественного языка необходимо учитывать иносказательность, возможность обучения, планирования и практическую сложность человеческих взаимоотношений. Таким образом, правильность распознавания речи обусловлена множеством лингвистических и экстралингвистических факторов.

Литература

[1] Бутенко Ю.И., Шостак И.В. Методологические аспекты автоматического распознавания речи на основе многомерной статистической теории. *Нейрокомпьютеры: разработка, применение*, 2018, № 2, с. 23–33.

[2] Селиванова Е.А. Лингвистическая энциклопедия Полтава. М., Довкиля-К, 2010.

[3] Потапова Р.К., Потапов В.В. Речевая коммуникация. М., Языки славянских культур, 2012.

[4] Фролов А.В. Синтез и распознавание речи. Современные решения. М., Связь, 2003.

[5] Бутенко Ю.И. Использование триграмм при автоматическом распознавании речи. *Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация*, 2020, т. 18, № 3, с. 5–15.

[6] Пиотровский Р.Г. Моделирование фонологических систем и методы их сравнения. Ленинград, Наука, 1966.

[7] Бутенко Ю.И., Ермакова Ю.В. Типы контекстов при распознавании устной речи. *XVII Всерос. науч. конф. Нейрокомпьютеры и их применение*. М., МГППУ, 2019, с. 183–185.

[8] Косарев Ю.А., Ли И.В., Ронжин А.Л. и др. Обзор методов понимания речи и текста. *Труды СПИИРАН*, 2002, т. 2, № 1, с. 157–195.

[9] О понимании компьютерами текста. URL: https://habr.com/ru/post/126748/ (дата обращения: 15.08.2019).

[10] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. СПб., Питер, 2000.

Кокурина Наталья Владимировна — студентка факультета «Романо-германские языки», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Жуков Даниил Михайлович — студент факультета «Информатика и управление», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Бутенко Юлия Ивановна, кандидат технических наук, доцент кафедры «Романо-германские языки», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:
Кокурина Н.В., Жуков Д.М. Роль контекста при автоматическом распознавании устной речи. *Политехнический молодежный журнал*, 2022, № 04(69).
http://dx.doi.org/10.18698/2541-8009-2022-04-784

THE ROLE OF CONTEXT IN AUTOMATIC SPEECH RECOGNITION

N.V. Kokurina kokurina.natasha@gmail.com
D.M. Zhukov dmzhukov@outlook.com

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

The context as a factor affecting the results of automatic speech recognition is considered. The necessity of classifying contexts in automatic natural language processing is described. The structural and non-structural types of contexts are outlined, and the intonational context is highlighted. The lexical and syntactic contexts in which the meaning of a lexical unit depends on the nearest environment are considered. Each of the identified contexts influencing the quality of automatic speech recognition is illustrated with examples and comments. It is emphasized that semantic and situational contexts are of particular importance for automatic speech recognition. The cultural context is highlighted as the most challenging for automatic speech processing.

Keywords

Context, speech recognition, speech understanding, micro-context, macro-context, spoken language, lexical context, structural context, syntactic context, cultural context, acoustics

Received 22.10.2021

References

[1] Butenko Yu.I., Shostak I.V. Methodological aspects of automatic speech recognition on the basis of multivariate statistical theory. *Neyrokomp'yutery: razrabotka, primenenie* [Neurocomputers], 2018, no. 2, pp. 23–33 (in Russ.).

[2] Selivanova E.A. Lingvisticheskaya entsiklopediya Poltava [Poltava linguistic encyclopedia]. Moscow, Dovkilya-K Publ., 2010.

[3] Potapova R.K., Potapov V.V. Rechevaya kommunikatsiya [Verbal communication]. Moscow, Yazyki slavyanskikh kul'tur Publ., 2012 (in Russ.).

[4] Frolov A.V. Sintez i raspoznavanie rechi. Sovremennye resheniya [Synthesis and recognition of speech. Modern solutions]. Moscow, Svyaz' Publ., 2003 (in Russ.).

[5] Butenko Yu.I. Using trigrams for automatic speech recognition. *Vestnik NGU. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya* [NSU Vestnik. Series: Linguistics and Intercultural Communication], 2020, vol. 18, no. 3, pp. 5–15 (in Russ.).

[6] Piotrovskiy R.G. Modelirovanie fonologicheskikh sistem i metody ikh sravneniya [Simulation of phonologic systems and methods for their comparison]. Leningrad, Nauka Publ., 1966 (in Russ.).

[7] Butenko Yu.I., Ermakova Yu.V. [Types pf contexts at oral speech recognition]. XVII *Vseros. nauch. konf. Neyrokomp'yutery i ikh primenenie* [XVII Russ. Sci. Conf. Neurocomputers and Their Application]. Moscow, MGPPU, 2019, pp. 183–185 (in Russ.).

[8] Kosarev Yu.A., Li I.V., Ronzhin A.L. et al. Review of speech and text understanding methods. *Trudy SPIIRAN* [SPIIRAS Proceedings], 2002, vol. 2, no. 1, pp. 157–195 (in Russ.).

[9] O ponimanii komp'yuterami teksta [On understanding of text by computers] (in Russ.). URL: https://habr.com/ru/post/126748/ (accessed: 15.08.2019).

[10] Gavrilova T.A., Khoroshevskiy V.F. Bazy znaniy intellektual'nykh system [Knowledge bases of intelligent systems]. Sankt-Petersburg, Piter Publ., 2000 (in Russ.).

Kokurina N.V. — Student, Department of Romance and Germanic Languages, Bauman Moscow State Technical University, Moscow, Russian Federation.

Zhukov D.M. — Student, Department of Informatics and Management, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Butenko Iu.I., Cand. Sc. (Eng.), Assoc. Professor, Department of Romance and Germanic Languages, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:
Kokurina N.V., Zhukov D.M. The role of context in automatic speech recognition. *Politekhnicheskiy molodezhnyy zhurnal* [Politechnical student journal], 2022, no. 04(69). http://dx.doi.org/10.18698/2541-8009-2022-04-784.html (in Russ.).