

КОНТЕКСТНОЕ РАСПОЗНАВАНИЕ СЛИТНОЙ РЕЧИ

Выхованец В.С., Ду Цзяньмин, Назарова С.И.

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва

МГТУ им. Н.Э. Баумана, г. Москва,

valery@vykhovanets.ru, forsola.du@gmail.com, nazarova_svetlana92@mail.ru

Ключевые слова: автоматическое распознавание речи, сегментация речи, фонетические грамматики, речевой протоязык, контекстное распознавание, разнесенный грамматический разбор, семантика и прагматика языка-речи.

Введение

Распознавание речи – область речевых технологий, включающая целый ряд достаточно обособленных направлений, каждое из которых ориентировано на решение конкретных прикладных проблем, требующих отдельной теоретической и практической проработки.

Одно из таких направлений – автоматическое распознавание речи, или преобразование речевого сигнала в текст. Выделяют несколько задач автоматического распознавания речи. Это распознавание отдельных команд для управления устройствами и приложениями, распознавание отдельных фраз в системах массового обслуживания, поиск ключевых слов в слитной речи в системах мониторинга и распознавание слитной речи на большом словаре [1].

Несмотря на то, что распознавание отдельных команд, фраз и ключевых слов имеет эффективную практическую реализацию, задача достоверного распознавания слитной речи пока не решена полностью. Проблемы возникают на всех стадиях обработки слитной речи: на стадии предварительной обработки речевых сигналов, при принятии решения «речь-пауза» и «вокализованный-невокализованный», на стадии сегментации сигнала на значимые фрагменты, при распознавании фонем, при сопоставлении последовательности фонем словам из словаря, при выявлении и расстановке ударений, на стадии учета грамматических правил словообразования и построения предложений, и т.п. [2].

Другой отличительной особенностью применяемых решений в распознавания речи является необходимость разделения процессов обработки на два этапа – на этап обучения и, соответственно, на этап распознавания [3]. Однако вариативность многих факторов речи зачастую не удается заранее учесть на этапе обучения.

Для оценки эффективности автоматического распознавания речи применяют множество показателей, одними из которых является точность распознавания и скорость обработки речевого сигнала. Но русский язык имеет относительно высокий уровень флективности (изменчивости). По этой причине пословное распознавание на большом словаре оказывается практически невозможным из-за низкой скорости обработки.

Ключевыми проблемами автоматического распознавания речи является ее высокая вариативность, возникающая из-за различия дикторов и изменчивости окружающей акустической обстановки, а также низкая достоверность распознавания, являющаяся следствием высокой флективности естественного языка-речи.

Причины неудач кроются в том, что речь, зафиксированная в форме звука, многоаспектна, а ее транскрипция в текстовую форму настолько обедняет звуковой поток, что становится невозможным надежное решение даже простых прикладных задач. Другая группа причин низкой надежности распознавания связана с тем, что эта ненадежность является объективным фактором речевого способа коммуникации: при восприятии речи человек тоже не все распознает надежно. Однако воспринимая речь, человек соотносит сказанное с действительностью, со своими знаниями о ней, со своим опытом, благодаря чему происходит восстановление пропущенных фрагментов. В процессе восприятия речи человек активен, выдвигает гипотезы относительно содержания фрагментов речи и осуществляет смысловые замены.

Настоящая статья посвящена технологии контекстного распознавания речи, которая основана на использовании фонетических грамматик естественных языков и баз знаний, описывающих предметные области устного дискурса. Учитывая поисковый характер выполненных исследований, задача достижения высокой скорости обработки речевого сигнала не ставилась.

Основное внимание было уделено на получение высокой надежности распознавания и достоверной интерпретации различных форм устного дискурса.

Отложенная слоговая сегментация

Сегментация потока устной речи на значимые единицы (предложения, паузы, слова, фонемы, аллофоны) оказалась довольно затруднительной. Выделяемые из этого потока фрагменты далеко не всегда обладают признаками синтаксических единиц, свойственных письменной речи. Установлено, что речевая коммуникация предполагается не только декодирование языковых единиц, но одновременное раскрытие их смыслов, т.е. сегментация речи через ее идентификацию [4], а также учета фоновых знаний, ситуации общения и структуры дискурса в самом процессе декодирования [5]. По этой причине транскрипция речевого потока в текстовую форму видится малоперспективной.

Хотя в формантной картине гласных в различных видах речи наблюдается определенная вариативность [6], сегментация слитной речи на основе выделения гласных звуков и с последующим построением на их основе фонетических слогов видится наиболее надежной. Фонетический слог – гласный или сочетание гласного с одним или несколькими согласными, произносимые одним выдохательным толчком речевого аппарата. Выделение фрагментов слитной речи, которые соответствуют фонетическим слогам, будем производить с учетом контекста распознавания, призванного учесть фоновые знания о предметной области дискурса и конкретную ситуацию речевого общения.

Контекст распознавания фонетического слога – это состояние фонетического анализатора, при котором гласный звук по-разному может быть соотнесен с образующим его слогом. При этом возможно присоединение к гласному звуку как предшествующих фрагментов речи, так и последующих, вплоть до граничащих с ним другими гласными звуками. В этом случае принятие решения о слоговой сегментации осуществляется на поздних этапах распознавания, когда к этому процессу присоединяется фонетическая грамматика языка и база знаний предметной области.

Фонетические грамматики

Грамматика языка – описание синтаксического строя языка, заключающегося в фиксации закономерностей построения его правильных и осмысленных речевых отрезков. Фонетика, как правило, не включается в грамматику и противопоставляется ей. Учитывая имеющий место быть вариативный характер сегментации и интерпретации отрезков слитной речи, будем использовать фонетическую грамматику языка, элементарными единицами которой являются не слова и словосочетания, а фонетические слоги, задаваемые фрагментами образцовых речевых сигналов.

Фонетическая грамматика выражает правила построения правильных речевых отрезков с учетом как синтаксических, так и фонетических правил языка. Эти правила во всей своей совокупности определяют некоторый язык, рассматриваемый как множество тех речевых отрезков, которые могут быть распознаны и проинтерпретированы без грамматических нарушений.

Слог (терм) определим как элементарную фонетическую единицу языка, состоящую из последовательности звуков, образующих фонетический слог. Словом (лексемой) будем называть элементарную синтаксическую единицу языка, представленную одним или несколькими слогами. Слово – это такие подмножества слогов языка, которым может быть приписан (сопоставлен) некоторый элементарный смысл. В свою очередь речью (текстом) будем называть последовательность слов, представленную некоторым речевым отрезком и предназначенную для выражения более сложного смысла. Речь, в отличие от слов, предполагает свое деление на смысловые части, в то время как слово такого деления не допускает. Здесь используется такая универсалия всех засвидетельствованных языков как принцип композиционности, открытый Г. Фреге [7]. Принцип композиционности утверждает, что семантика целого определяется через семантику его частей, а ход (этапность) интерпретации определяется его структурой.

Таким образом, каждому нетерминальному знаку фонетической грамматики сопоставим некоторое нетерминальное понятие описываемого языка-речи. Тогда продукции грамматики

могут быть проинтерпретированы как фонетические и синтаксические правила выражения понятий предметной области дискурса в речи.

Речевой протоязык

За внешним многообразием естественных языков, которых насчитывается около шести тысяч, видны общие свойства коммуникативной системы человека. Так, во всех языках имеется деление звучащей речи на слоги – соответственно, противопоставлены гласные (слоγοобразующие) и согласные (неслоγοобразующие) звуки. Во всех языках представлено различие между номинацией (называнием), предикацией (суждением) и дискурсом (умозаключением). Все языки имеют средства для выражения прагматических компонентов высказываний. Во всех языках есть слова, предложения, а между отдельными составляющими синтаксической структуры предложений могут быть усмотрены закономерные корреляции. Все это дает возможность говорить о наличии некоторой универсальной грамматики – общей врожденной подосновы всех засвидетельствованных и потенциально возможных языков, или протоязыка [8]. Однако в настоящей работе ставится задача не анализа исторического пути развития и становления естественных языков, а задача поиска и выражения наиболее общих и значимых свойств современного языка-речи. Формализация этих свойств позволит в наиболее простой и естественной форме описать синтаксический и фонетический строй некоторого ограниченного языка для заданной предметной области. Только на основе такого описания может быть решена проблема автоматического распознавания слитной речи.

Для описания фонетических грамматик воспользуемся протоязыком. В протоязык включим средства для выражения наиболее устойчивых артефактов, лежащих в основе как синтаксического, так и фонетического строя языка. Семантическую роль протоязыка ограничим реализацией аксиомы, позволяющей ввести в использование первичные семантические категории, чрез которые и посредством которых осуществляется интерпретация распознанных речевых отрезков. Формальный математический аппарат протоязыка – контекстно-свободные грамматики [9], позволяющие описывать грамматические правила в виде последовательностей терминальных и нетерминальных знаков, а также группировать эти правила по выражаемому ими нетерминальному знаку.

Для каждой предметной области будем создавать свой проблемный язык-речь, отражающий ее понятийную структуру (онтологию). Для этого выявленные в процессе анализа предметной области понятия включим во множество понятий проблемного языка, а способы выражения понятий положим в основу его фонетики и синтаксиса [10]. Описание понятийной структуры, фонетики и синтаксиса языка-речи выполним на протоязыке, имеющем фиксированные фонетику, синтаксис и семантику, которые минимально достаточны для определения фонетики, синтаксиса и семантики любого проблемного языка

Контекстное распознавание

После описания фонетической грамматики интерпретация любого речевого фрагмента осуществляется в процессе контекстного распознавания предъявленного этого фрагмента с фонетическими единицами – образцовыми речевыми фрагментами, синтаксическими единицами, входящими в синтаксические правила языка, и синтаксическими правилами, представленными продукциями контекстно-свободной грамматики. При этом анализу подвергаются не все имеющиеся элементарные фонетические единицы, а только те, которые предусмотрены текущим синтаксическим правилом и могут быть выявлены во входном речевом потоке.

Под контекстным распознаванием будем понимать определение фонетического значения слога, а также синтаксического и семантического значений слова по их месту в слитной речи. В общем случае один и то же слог может быть сопоставлен различным словам, т.е. одна и та же последовательность слогов может служить носителем различных смыслов. Следовательно, при контекстной интерпретации фонетическое значение слога и семантическое значение слова определяется текущим контекстом.

Так как известные методы грамматического разбора не могут быть использованы в разрабатываемой технологии по причине отсутствия контекстных ограничений на формы выражения понятий, то для реализации фонетического, лексического, синтаксического и семантического

анализа речи применим метод разнесенного грамматического разбора [11]. Разнесенный грамматический разбор позволяет эффективно выполнить анализ слитной речи, порожденной контекстными и неоднозначными фонетическими грамматиками.

Семантика и прагматика

Описание семантики осуществим методом семантической индукции, заключающимся в использовании семантических категорий, которые определяются по мере необходимости, в процессе описания проблемного языка и средствами этого языка [12]. Базу индукции, или первичные семантические категории, объявим с помощью аксиомы протоязыка. На содержательном уровне это выглядит как использование создаваемого проблемного языка для описания своей семантики. Последнее выполнимо благодаря возможности пополнения множества базовых семантических категорий и наличия механизма определения новых семантических категорий на основе имеющихся.

Для задания понятийной структуры предусмотрим в протоязыке выразительные средства для описания способов абстрагирования понятий, а именно: типизации-конкретизации, обобщения-специализации, агрегации-декомпозиции и ассоциации-индивидуализации [13]. Выделенные абстракции рассматриваются как умственные операции, необходимые и достаточные для мысленного выделения и превращения в отдельные понятия тех представлений, которые накоплены относительно формализуемой предметной области.

Так как понятийная структура проблемной области может быть общей для нескольких задач, решение которых может преследовать различные цели, в том числе и исключающие друг друга, при описании семантики предусмотрим возможность задания одной или нескольких прагматик, являющихся, по своей сути, различными контекстными интерпретациями одной и той же формы выражения понятия.

Заключение

В настоящей работе рассмотрены накопившиеся проблемы в области распознавания слитной речи и предложен подход к их решению на основе отложенной слоговой сегментации, использования фонетической грамматики предметной области и соответствующего ей проблемного языка-речи, описания онтологии, фонетики, синтаксиса, семантики и прагматики этого языка.

Литература

1. *Потапова Р.К.* Речевое управление роботом. – М., Комкнига, 2005. – 328 с.
2. *Шелепов В.Ю.* К проблеме пофонемного распознавания // Искусственный интеллект. – 2005, № 4. – С. 662-668.
3. *Петровский А.А.* Методы построения устройств распознавания речи на базе гибрида «нейронная сеть» – «скрытая Марковская модель» // Нейрокомпьютеры: разработка и применение. – 2002, № 12. – С. 26-36.
4. *Венцов А.В., Касевич В.Б.* Проблемы восприятия речи. – М.: Едиториал УРСС, 2003. – 240 с.
5. *Андреева С.В.* Речевые единицы русской речи: Система, зоны употребления, функции. – М.: УРСС, 2006. – 192 с.
6. *Сорокин В.Н., Цыплихин А.И.* Сегментация и распознавание гласных / Информационные процессы. 2004, т. 4, №2. – С. 202-220.
7. *Демьянков В.З.* Предикаты и концепция семантической интерпретации // Изв. АН СССР. Сер. Лит. и языка. – Т. 39, № 4. – 1980.
8. *Барулин А.Н.* Основания семиотики. Знаки, знаковые системы, коммуникация. Ч. 1. Базовые понятия. Эволюционная теория происхождения языка. – М.: Спорт и культура, 2002.
9. *Гладкий А.В.* Формальные грамматики и языки. М.: Наука, 1973.
10. *Выхованец В.С.* Контекстная технология программирования // Труды IV-й Международной конференции «Параллельные вычисления и задачи управления». М.: Институт проблем управления, 2008. С. 1288-1327.
11. *Выхованец В.С.* Разнесенный грамматический разбор // Проблемы управления. 2006. № 3. С. 32-43.
12. *Выхованец В.С.* Описание семантики контекстно-свободных языков методом математической индукции // Научно-техническая информация. Сер. 2: Информационные процессы и системы. 2008. № 7. С. 6-14.

Материалы VII-й Международной конференции «Управление развитием крупномасштабных систем» (MLSD'2013). – Том II. – М.: ИПУ РАН, 2013. – С. 312-315.

13. *Выхованец В.С.* Исчисление понятий // Труды VII-й Международной конференции «Когнитивный анализ и управление развитием ситуаций» (CASC'2007, Москва). М.: Институт проблем управления, 2007. С. 31-35.