

§ МЕТОДЫ, ЯЗЫКИ И МОДЕЛИ ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ

Денисенко В.А., Ксалов А.М., Гошокова Ф.М.

МЕТОДИКА РАЗБОРА АГГЛЮТИНАТИВНЫХ ЯЗЫКОВ ДЛЯ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ИНТЕРФЕЙСА В МУЛЬТИАГЕНТНЫХ СИСТЕМАХ И СИСТЕМАХ ОСНОВАННЫХ НА ЗНАНИЯХ¹

Аннотация. Одна из основных задач искусственного интеллекта - сделать общение человека с машиной проще, в частности, предоставить возможность общения на естественном языке. Реализации этой цели служит создание естественно-языковых интерфейсов - программ, которые воспринимают текстовое естественно-языковое сообщение и либо переводят его в некоторый формальный язык, либо выполняют некоторое действие в соответствии с принятым сообщением [1]. Для того, чтобы система могла понять смысл текста, она должна разобрать каждое слово предложения а также соотношение этих слов внутри предложения. Разбор слов в агглютинативных языках на первый взгляд кажется довольно сложной задачей, поскольку такие языки характеризуются богатством словообразовательных и словоизменительных аффиксов. Однако аффиксы в агглютинативных языках выражают только одно значение, что приводит к нанизыванию аффиксов при выражении разных грамматических значений. Другая особенность аффиксов - они могут примыкать к основе с разных сторон. Также в агглютинативных языках обычно отсутствуют разнообразные типы склонения и спряжения [2]. Одним из агглютинативных языков является башкирский язык, для которого создана программа автоматического анализа башкирской морфологии *Bashmorph*. Эта программа умеет определять значение введенного в него слова, а также в форме какого лица (числа, падежа, наклонения и т.д.) оно стоит. На основе такой программы можно строить частотные словари, автоматические переводчики, поисковые системы и многое другое [3]. Технический прогресс ставит новые задачи, вводит новые термины и изменяет среду обитание человека. Все эти изменения отражает в себе язык, являющийся средством общения людей между собой, а в будущем – человека с машиной. Создание естественно-языковых интерфейсов для агглютинативных языков является важной задачей для сохранения и дальнейшего развития таких языков.

Ключевые слова: парсер, грамматика, разбор, агглютинативный язык, мультиагентная система, естественно-языковой интерфейс, *explainer*, агент, человеко-машинное взаимодействие, система коммуникаций.

¹ Работа выполнена при поддержке грантов РФФИ 12-07-01003-а, 13-07-01003-а

Введение

Одна из основных задач искусственного интеллекта — сделать общение человека с машиной проще, в частности, предоставить возможность общения на естественном языке. Реализации этой цели служит создание естественно-языковых интерфейсов — программ, которые воспринимают текстовое естественно-языковое сообщение и либо переводят его в некоторый формальный язык, либо выполняют некоторое действие в соответствии с принятым сообщением [1]. Для того, чтобы система могла понять смысл текста, она должна разобрать каждое слово предложения а также соотношение этих слов внутри предложения. Разбор слов в агглютинативных языках на первый взгляд кажется довольно сложной задачей, поскольку такие языки характеризуются богатством словообразовательных и словоизменительных аффиксов. Однако аффиксы в агглютинативных языках выражают только одно значение, что приводит к нанизыванию аффиксов при выражении разных грамматических значений. Другая особенность аффиксов — они могут примыкать к основе с разных сторон. Также в агглютинативных языках обычно отсутствуют разнообразные типы склонения и спряжения [2].

Одним из агглютинативных языков является башкирский язык, для которого создана программа автоматического анализа башкирской морфологии Bashmorph. Эта программа умеет определять значение введенного в него слова, а также в форме какого лица (числа, падежа, наклонения и т.д.) оно стоит. На основе такой программы можно строить частотные словари, автоматические переводчики, поисковые системы и многое другое [3].

Технический прогресс ставит новые задачи, вводит новые термины и изменяет среду обитание человека. Все эти изменения отражает в себе язык, являющийся средством общения людей между собой, а в будущем — человека с машиной. Создание естественно-языковых интерфейсов для агглютинативных языков является важной задачей для сохранения и дальнейшего развития таких языков.

Целью данного исследования является разработка алгоритма и программного продукта (парсера) для морфологического анализа слов агглютинативного языка применяемого для создания модуля естественно-языкового интерфейса.

Целью данной работы является разработка системы естественно-языкового интерфейса на основе системы разбора агглютинативного языка, которая может использоваться в мультиагентных системах и системах основанных на знаниях.

Задачи исследования сформулированных следующим образом:

- изучить технологии представления агглютинативного языка в компьютерной лингвистике;
- реализовать грамматики, описывающие различные части речи агглютинативного языка;
- разработать алгоритм разбора грамматик, описывающих агглютинативный язык.

Система разбора агглютинативных языков

Словоизменительный принцип в агглютинативных языках, подразумевающий регулярный порядок присоединения аффиксов к основе, низкий процент грамматической омонимии, отсутствие инфиксов и прочих подобных случаев, затрудняющих реализацию автоматического распознавания грамматической формы слова, облегчают практическую реализацию программного средства для морфологического разбора слова [4].

Морфологический разбор слова — это разбор слова по значимым морфологическим единицам (морфемам). Для автоматического морфологического разбора слов можно использовать формальные грамматики.

Алгоритм разбора различных частей речи представлен следующей схемой: см. рис. 1.

В рамках исследования были рассмотрены особенности формирования глаголов (как самых сложных частей речи для анализа в агглютинативных языках) и имен существительных. А также были составлены формальные грамматики для них. Для проверки выводимости входного слова из некоторой грамматики

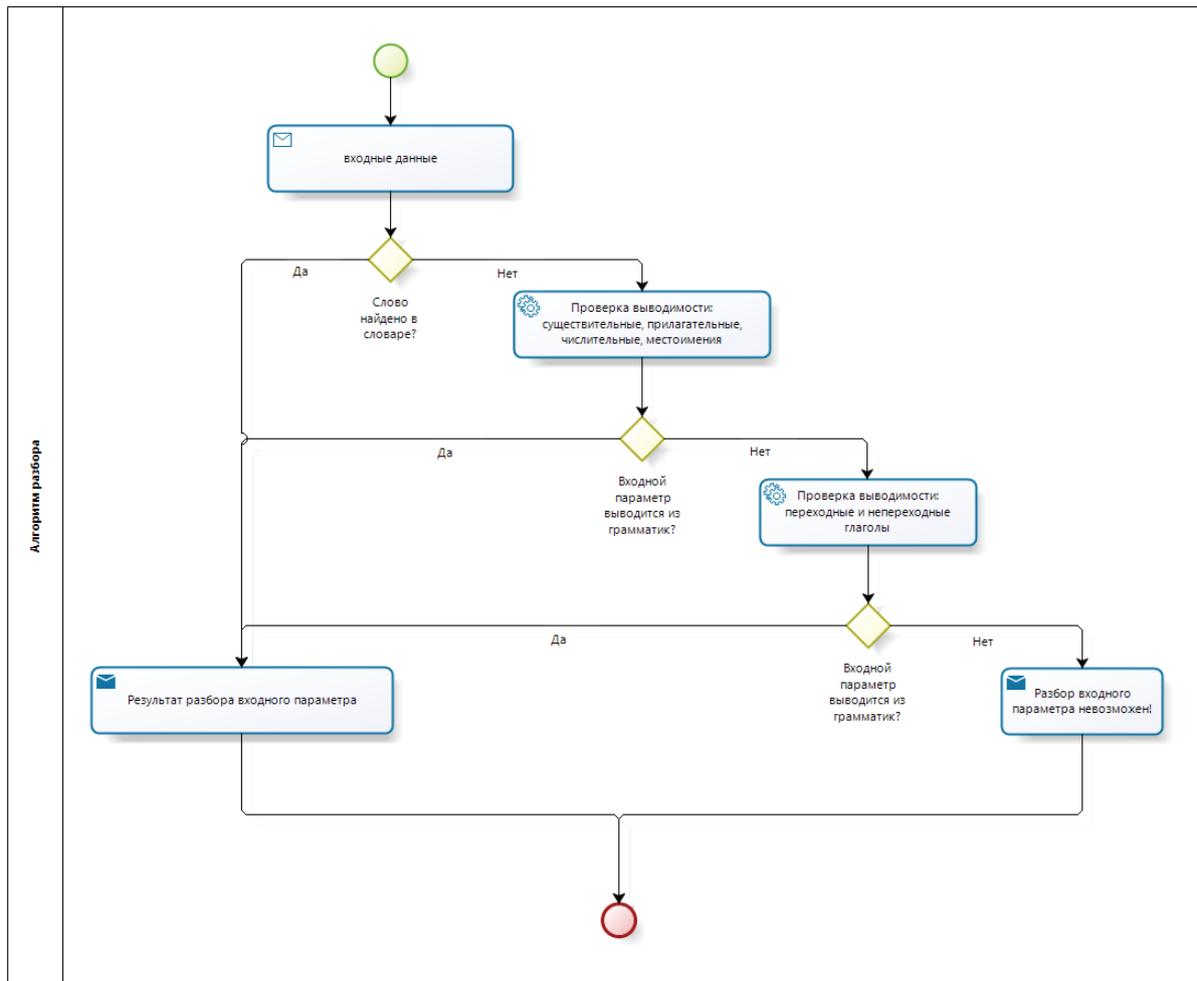


Рис. 1. Схема разбора агглютинативного языка

Таблица 1.

Фрагмент формальной грамматики

Имя существительное
S>KABTV КАВАНV
K>#
X>
A>g X
g>и ы
V>a b c d e f X
a>p
b>м
c>мкIэ
d>рауэ
e>кIэ
f>у

был разработан парсер на основе метода рекурсивного спуска. Метод рекурсивного спуска — алгоритм синтаксического анализа, реализуемый путём взаимного вызова парсящих процедур, соответствующих правилам контекстно-свободной грамматики, являющейся частным случаем формальной грамматики.

В таблице 1 представлен фрагмент формальной грамматики для имени существительного кабардино-черкесского языка, являющегося агглютинативным языком.

Используя формальную грамматику и имея набор правил для этой грамматики, морфологический анализатор (парсер) в автоматическом

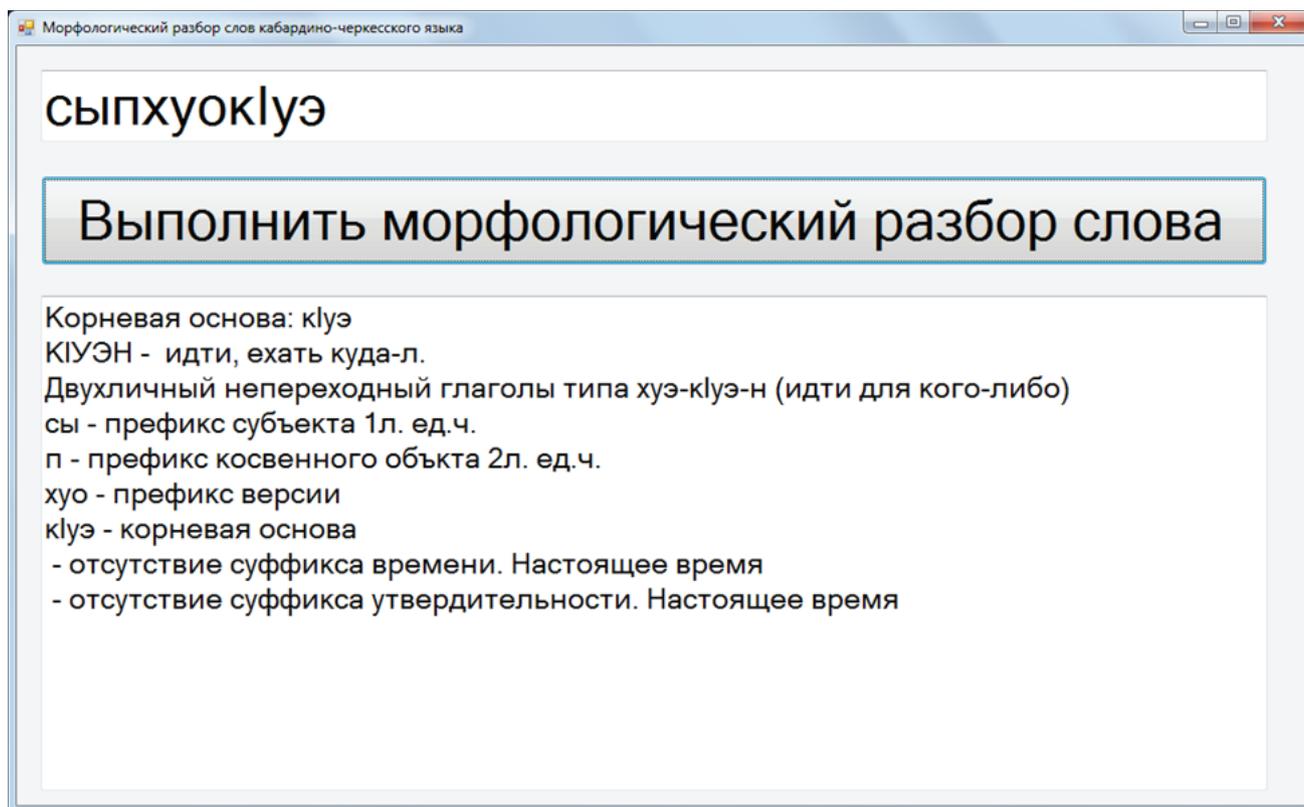


Рис.2. Пример работы парсера глаголов агглютинативного языка

режиме разбирает слова по морфемам, выдает инфинитную форму, число, падеж, наклонение и т.д. А благодаря подключенному словарю выдается и словарная карточка данного слова.

На рисунке 2 показана работа парсера глагола кабардино-черкесского языка.

Разработка модуля естественно-языкового интерфейса

Естественно-языковой интерфейс — разновидность пользовательского интерфейса, который принимает запросы на естественном языке, а также, возможно, использует естественный язык и для вывода информации. То есть по запросу пользователя система должна адекватно реагировать — либо выполнять какие-либо действия, запрашиваемые пользователем, либо выдавать интересующую его информацию. Поступающий запрос от пользователя будет анализироваться, каждое слово будет разбираться морфологическим анализатором, а затем естественно-языковой модуль должен будет сформировать сообщение на основе сделанного разбора запроса.

Конечно, использование естественного языка очень удобно для пользователя и имеет много преимуществ, таких как простота задания запросов, большая скорость задания запросов, минимальная предварительная подготовка и т.д. Но такой способ имеет и свои недостатки, главным из которых является неоднозначность естественного языка. Несмотря на это очевидно, что естественно-языковой интерфейс войдет в нашу жизнь в недалеком будущем.

Выводы

Разработанные в рамках исследования методы разбора и анализа слов агглютинативного языка могут быть использованы для различных задач, в том числе и для создания модуля естественно-языкового интерфейса.

Библиография

1. Шишкин Д. В. Система построения естественно-языковых интерфейсов, ориентированная на классификацию текстов в иерархически структурированной предметной области. Материалы международной конференции молодых ученых по математическому моделированию и информационным технологиям, 2002 года, Новосибирск
2. Зенков Г. С., Сапожникова И. А. Введение в языкознание. Учебное пособие для студентов дистанционного обучения КГНУ. Бишкек: ИИМОПКГНУ, 1998.— 218 с.
3. Документация к морфологическому анализатору башкирского языка bashmorph: <http://lcpb.bashedu.ru/index.php?go=parserdoc>
4. Орехов Б. В., Слободян Е. А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 2010 г.) / отв. ред. В. А. Баранов.— Уфа; Ижевск: Вагант, 2010.— С. 167–171.
5. Денисенко В. А., Ксалов А. М. Разработка парсера глаголов агглютинативного языка для системы коммуникации мультиагентного коллектива // Известия КБНЦ РАН № 2 (46) 2012
6. Кабардино-черкесский язык в 2-х тт / гл. ред. М. А. Кумахов.— Нальчик: Эль-Фа, 2006.
7. В. А. Денисенко, А. М. Ксалов, Ф. М. Гошокова. Алгоритм вывода неопределенной формы глагола агглютинативного языка на основе грамматик. // Известия КБНЦ РАН № 5 (49) 2012
8. Денисенко В. А., Ксалов А. М. Парсер глаголов агглютинативного языка для системы коммуникации мультиагентного коллектива. Материалы Второй международной конференции «Автоматизация управления и интеллектуальные системы и среды», Красная поляна, 15–23 декабря 2011.

References (transliterated)

1. Shishkin D. V. Sistema postroeniya estestvenno-yazykovykh interfeisov, orientirovannaya na klassifikatsiyu tekstov v ierarkhicheski strukturirovannoi predmetnoi oblasti. Materialy mezhdunarodnoi konferentsii molodykh uchenykh po matematicheskomu modelirovaniyu i informatsionnym tekhnologiyam, 2002 goda, Novosibirsk
2. Zenkov G. S., Sapozhnikova I. A. Vvedenie v yazykoznanie. Uchebnoe posobie dlya studentov distantsionnogo obucheniya KGNU. Bishkek: IIMOPKGNU, 1998.— 218 s.
3. Dokumentatsiya k morfologicheskomu analizatoru bashkirskogo yazyka bashmorph: <http://lcpb.bashedu.ru/index.php?go=parserdoc>
4. Orekhov B. V., Slobodyan E. A. Problemy avtomaticheskoi morfologii agglyutinativnykh yazykov i parser bashkirskogo yazyka // Informatsionnye tekhnologii i pis'mennoe nasledie: materialy mezhdunarodnoi nauchnoi konferentsii (Ufa, 2010 g.) / отв. red. V. A. Baranov.— Ufa; Izhevsk: Vagant, 2010.— S. 167–171.
5. Denisenko V. A., Ksalov A. M. Razrabotka parseera glagolov agglyutinativnogo yazyka dlya sistemy kommunikatsii mul'tiagentnogo kollektiva // Izvestiya KBNTs RAN № 2 (46) 2012
6. Kabardino-cherkesskii yazyk v 2-kh tt / gl. red. M. A. Kumakhov.— Nal'chik: El'-Fa, 2006.
7. V. A. Denisenko, A. M. Ksalov, F. M. Goshokova. Algoritm vyvoda neopredelennoi formy glagola agglyutinativnogo yazyka na osnove grammatik. // Izvestiya KBNTs RAN № 5 (49) 2012
8. Denisenko V. A., Ksalov A. M. Parser glagolov agglyutinativnogo yazyka dlya sistemy kommunikatsii mul'tiagentnogo kollektiva. Materialy Vtoroi mezhdunarodnoi konferentsii «Avtomatizatsiya upravleniya i intellektual'nye sistemy i sredy», Krasnaya polyana, 15–23 dekabrya 2011.