

## **АННОТАЦИЯ**

**Диссертационной работы докторанта PhD  
специальности 8D06103 – “Management information systems”**

**Бекарыстанкызы Акбаян**

**на тему: “Разработка интегральной системы для автоматического  
распознавания речи на агглютинативных языках”**

**Актуальность темы исследования.** Системы автоматического распознавания речи (САРР) в наши дни очень широко используются в разных сферах человеческой жизни для облегчения взаимодействия человека с компьютерными системами и разными приложениями. Например, умные помощники, системы умного дома, системы автоматического создания субтитров позволяют управлять компьютерные системы и устройства без прикосновения, даже можно сказать издалека. Более того, системы автоматического распознавания речи облегчают использование электронных устройств для людей, которые имеют проблемы со здоровьем. Например, группа ученых изучают способы создания системы автоматического распознавания речи для людей с дизартрией. Дизартрия — это тип дефектов мышц, ответственных за артикуляцию. Разработка САРР для таких людей помогают им не только взаимодействовать с цифровыми системами, а также общаться с другими людьми. Следующий полезный пример внедрения САРР, это оценка степени потери слуха. Этот тип системы может оценить уровень нарушения слуха с помощью оценки ответа на заданные вопросы. Но эти возможности доступны только для людей, которые владеют широко используемые языки, такие как Английский, Китайский, Русский и т.д. Разработка САРР для мало ресурсных языков до сих пор требует больших усилий, таких как сбор данных и их обработка, тестирование очень хорошо знакомых архитектур, также приспособление так называемых state-of-the art архитектур для конкретных языков или для групп языков. К Турецкой группе агглютинативных языков помимо Казахского языка относятся много мало ресурсных языков. Помимо недостатка данных для обучения, агглютинативные языки имеют другие проблемы, которые перечислены ниже.

Разработка систем ASR для агглютинативных языков представляет собой сложный процесс из-за их морфологической сложности и богатства грамматических форм в этих языках. Согласно этому, разработка и подгонка систем АРР для агглютинативных языков требуют дополнительных исследований и специфических подходов. Ниже приведен список нескольких

проблем, с которыми могут столкнуться системы АРР для агглютинативных языков:

- Анализ и разделение морфем. Морфемы в агглютинативных языках могут соединяться и быть сложными, что затрудняет процесс расщепления и анализа морфем при распознавании речи.
- Разнообразие правил словообразования: в агглютинативных языках обычно существуют различные правила словообразования, которые определяют, как соединять аффиксы с корнем слова. Это также требует сложных моделей и правил для обработки этих правил в системах АРР.

Существует огромное количество исследований, посвященных разработке конкретных подходов и моделей получения надежных САРР для агглютинативных языков. Некоторые авторы предлагают языковую модель, основанную на морфемах, где под морфемами понимается любая приставка, корень или суффикс в слове. В результате авторы получили систему автоматического распознавания речи с большим словарным запасом. Одно исследование изучает производительность системы СТС на основе трансформера, которая зависит от контекста и обучается с использованием фрагментов слов, взятых в качестве обучающих единиц. Авторы отмечают эффективность своего метода не только для английского и немецкого языков, но и для одного из агглютинативных языков - турецкого. В следующем исследовании изучалось применение архитектуры трансформер для устранения морфологической неоднозначности с использованием турецкого языка. Это средство устранения неоднозначности можно использовать в любой задаче НЛП, и распознавание речи здесь не является исключением. Архитектура трансформатора хорошо зарекомендовала себя и для другого агглютинативного языка — хинди. Здесь архитектура преобразователя вместе с коннекционной временной классификацией (СТС) и языковой моделью показала самый низкий уровень ошибок для языка хинди: 3,2%. Еще одним примером использования архитектуры трансформера для разработки САРР для агглютинативного языка является финский язык. Здесь автор сравнил производительность архитектуры Transformer-XL с LSTM и пришел к выводу, что улучшение значения «недоумения» (perplexity) достигло 29%, а частота ошибок в словах снизилась до 3% для финского языка.

В следующей статье авторы утверждают, что существуют наиболее широко используемые и эффективные архитектуры сквозного автоматического распознавания речи: коннекционная временная классификация и механизм внимания. Также в этой работе упоминается недостаточность ресурсов транскрибированных пар аудио-текст для агглютинативных языков для

обучения с целью разработки надежных систем автоматического распознавания речи.

На основании вышеперечисленных исследований для агглютинативных языков было отмечено, что расширение словаря и архитектура трансформер являются наиболее эффективными подходами для разработки сквозных систем автоматического распознавания речи для агглютинативных языков. Более того, отсутствие данных для обучения, единые морфологические правила и схожести звучания языков тюркской семьи агглютинативных языков послужило основой для проведения экспериментов по объединению, таких как трансферное обучение и многоязычное обучение для этих языков.

**Цель диссертационной работы.** Настоящая диссертация разработана с целью изучения путей повышения производительности ASR для агглютинативных языков на примере языков тюркского семейства.

#### **Задачи исследования**

- 1) Анализ существующих методов и подходов для автоматического распознавания речи для общего случая и для агглютинативных языков.
- 2) Разработка и увеличение корпуса данных для агглютинативных языков тюркского семейства.
- 3) Разработка моделей и методов для автоматического распознавания речи на агглютинативных языках.
- 4) Разработка системы для автоматического распознавания речи на агглютинативных языках.

**Объект исследования.** Современные подходы и методы автоматического распознавания речи, особенно такие методы как мультязычное обучение и трансферное обучение.

**Предмет исследования.** Агглютинативные языки Тюркского семейства языков, методы машинного обучения, а именно нейронные сети для автоматического распознавания речи, сверточные нейронные сети, методы улучшения распознавания для языков с критическим низким уровнем, вложения слов и векторы для представления слов, демонстративный Телебот возможностей обученной системы автоматического распознавания речи и приложение, переводящий содержимое аудиофайлов в текст.

**Методы исследования.** Методы машинного обучения, методы и технологии автоматического распознавания речи, методы обработки естественного языка, математическая статистика и теория вероятностей.

**Научная новизна.** Диссертационная работа выдвигает научные и практическую новизны, которые были применены для практических заданиях, особенно для улучшения интегральных систем автоматического распознавания речи для агглютинативных языков с фокусом на казахский язык. Предложенные подходы также могут быть полезны и для других языков. Также

был внесен значимый вклад в увеличение размера размеченных данных для Казахского языка. Главные положительные результаты, полученные в ходе исследований перечислены ниже:

- 1) Был разработан корпус данных для агглютинативных языков.
- 2) Были разработаны эффективные модели для распознавания мало-ресурсных агглютинативных языков тюркского семейства: трансферный метод, мультиязычное обучение, расширенная языковая модель.
- 3) Система для автоматического распознавания речи для агглютинативных языков.

**Теоретическое и практическое значение работы.** Теоретическая значимость исследования заключается в том, что оно предлагает возможность улучшения производительности АРР, улучшая только языковую модель с внешним «большим текстом», и показывает возможность улучшения производительности для всех языков из одного семейства языков, включенных в многоязычное обучение, трансферное обучение. Возможность применения всех упомянутых теоретических положений для обучения САРР для агглютинативных языков тюркской семьи показывает практическую значимость данной диссертации. Более того, алгоритм предобработки текста при сборе текста может применяться для решения широкого круга задач обработки текста. Данные пары аудио-текст, собранные в ходе исследования, могут быть использованы в различных задачах обработки речи.

**Основное положение, выносимое на защиту**

- 1) Разработан корпус данных для агглютинативных языков.
- 2) Методы улучшения автоматического распознавания речи на мало-ресурсных агглютинативных языках.
- 3) Была получена система автоматического распознавания речи для агглютинативных языков.

**Степень достоверности и апробации результатов.** Исследования и их результаты в рамках темы диссертационной работы были представлены и обсуждены в разных конференциях и семинарах, и некоторые были опубликованы. Также, автор была награждена сертификатами, как спикер семинара и как самая лучшая презентация доклада:

- 1) О. Мамырбайев, D. Oralbekova, A. Kudyrbekova, T. Turdalykyzy and A. Bekarystankyzy, "End-to-End Model Based on RNN-T for Kazakh Speech Recognition," 3rd International Conference on Computer Communication and the Internet (ICCCI) (25-27 июня 2021 г., Токио).
- 2) Сертификат спикера семинара, на тему "Improved Speech Recognition for Agglutinative languages", Coimbra Institute of Engineering (ISEC), (21 апреля 2023 г., Коимбра, португалия).

- 3) Сертификат за самую лучшую презентацию доклада на тему “Improve Automatic Speech Recognition for Kazakh Language using Extended Language Model”, “ACeSYRI Young Researchers School” (5-10 июня 2023 г., Алматы, Казахстан).\_
- 4) A. Bekarystankyzy, O. Mamyrbayev, “IMPROVE AUTOMATIC SPEECH RECOGNITION FOR KAZAKH LANGUAGE USING EXTENDED LANGUAGE MODEL”, 21-научная конференция, (20-21 апреля 2023г., Рига, Латвия).
- 5) Automatic Speech Recognition Improvement for Kazakh Language with Enhanced Language Model // Recent Challenges in Intelligent Information and Database systems. ACIIDS 2023. Part of The Communications in Computer and Information Science book series. – 2023. - Vol. 1, - P.538-545 (Springer, Cham).

**Личный вклад исследователя.** PhD-студент самостоятельно выполнила и решила задачи диссертационной работы. Автор самостоятельно реализовала интегральные модели распознавания речи для Казахского и Агглютинативных языков. Внесла свой личный вклад в увеличение размеченного корпуса казахского языка, особенное во включении данных, собранных в реальных условиях. Спроектировала и выполнила эксперименты и их оценки над существующими и усовершенствованными моделями.

**Связь темы диссертации с планами научно-исследовательской работы.**

Исследования в рамках темы диссертационной работы были проведены в пределах грантового проекта: «Разработка интегральной системы автоматического распознавания речи для агглютинативных языков» (2020-2022, государственный регистрационный номер: 0120PK00344) в Институте информационных и вычислительных технологий Комитета науки МОН РК.

**Публикация основных результатов диссертационного исследования.** В рамках темы исследования были опубликованы несколько статей, один из которых был опубликован в периодическом журнале с ненулевым импакт-фактором и индексируемый в базах данных Scopus и Web of Science, и три из которых были опубликованы в журналах, рекомендованных Комитетом по контролю в сфере образования и науки МОН РК.

- 1) M. Orken, A. Keylan, O. Dina, B. Akbayan and Z. Bagashar. (2022). Identifying the influence of transfer learning method in developing an end-to-end automatic speech recognition system with a low data level // Eastern-European Journal of Enterprise Technologies. - Vol. 1, № 115. -P. 84–92 // <https://doi.org/10.15587/1729-4061.2022.252801>(Scopus, percentile 34);
- 2) Bekarystankyzy A. and Mamyrbayev O. (2023). INTEGRATED AUTOMATIC SPEECH RECOGNITION SYSTEM FOR

AGGLUTINATIVE LANGUAGES // News of the National academy of sciences of the republic of Kazakhstan. - 2023. - Vol. 1, № 345. -P. 37-49 // <https://doi.org/10.32014/2022.2518-1726.167>

- 3) Bekarystankyzy A., Mamyrbayev O., Oralbekova D., Zhumazhanov B. (2023). Transfer learning for an integrated low-data automatic speech recognition system // Scientific and technical journal "Bulletin of the Almaty University of Power Engineering and Telecommunications". - 2023, -Vol. 1, №. 60. -P. 185-198 // [https://doi.org/10.51775/2790-0886\\_2023\\_60\\_1\\_185](https://doi.org/10.51775/2790-0886_2023_60_1_185)
- 4) Bekarystankyzy A. and Mamyrbayev O. (2023). End-to-end speech recognition systems for agglutinative languages // Scientific Journal of Astana IT University. - 2023. -Vol. 13. -P. 86-92 // DOI: 10.37943/13IMII7575
- 5) Авторское свидетельство "Программное обеспечение UniCodeKaz" № 38545 от 21.08.2023, Автор: Бекарыстанкызы А.
- 6) Авторское свидетельство "Система автоматического транскрибирования аудио файлов в текст " № 38833 от 31.08.2023, Авторы: Бекарыстанкызы А., Мамырбаев О., Дуйсенхан Б.

**Структура и объем диссертационной работы.** Диссертационная работа состоит из введения, 4 разделов, заключения, списка литературы из 163 наименований и 4 приложений. Работа изложена на 107 страницах и содержит 38 рисунков, 16 таблиц.

#### **Краткое описание диссертационного исследования**

Во **введении** представлены сведения об актуальности темы исследования, цели, вопросах исследования и задачах диссертации, научной новизне, теоретической и практической значимости исследования и методах исследования, использованных в диссертации.

В **первой** главе представлен всесторонний обзор соответствующих работ как в агглютинативных, так и в неагглютинативных языках. Первоначальное обсуждение в этой главе посвящено методологии изучения современных случаев. Далее рассматриваются первые примеры САРР, внедрение нейронных сетей в область обработки речи, современные модели САРР. Последние два подраздела этой главы посвящены исследованиям и достижениям агглютинативных языков и казахского языка.

Во второй главе, посвященной теоретическим основам распознавания речи, рассматриваются наиболее важные концепции агглютинативных языков, обработки естественного языка, автоматического распознавания речи, показатели производительности и модели распознавания речи, начиная с математического и акустического моделирования и заканчивая обсуждением основных подходов, такие как скрытые Марковские модели, модели гауссовой смеси, скрытые Марковские модели/искусственные нейронные сети, сквозные модели и типы сетевых блоков для моделей последовательностей, таких как

RNN, GRU и LSTM. В последнем подразделе описываются современные архитектуры, основанные на механизме внимания и коннекционной временной классификации: трансформер, конформер и бранчформер.

Третья глава посвящена экспериментам и их результатам. Эта глава начинается с процесса сбора данных. Далее идут разделы посвящённые мультязычному эксперименту для агглютинативных языков Тюркского семейства и имеющих одинаковые и созвучные буквы в алфавитах; увеличенной языковой модели полученной обучением «Большого текста» на архитектуре трансформер и используемой на стадии декодинга; трансферному обучению, проведенному для Казахского и Азербайжанского языков.

В **четвертой** главе обсуждаются результаты всех экспериментов, включенных в диссертационную работу, и их результаты, причем эти результаты сравниваются с результатами предыдущих исследований, аналогичных проведенным исследованиям.

В **заключении** суммируются основные результаты исследований, включенных в диссертационную работу, и содержится информация о будущей работе, которая будет проведена впоследствии.