



3. T.F. Cootes and C.J. Taylor and D.H. Cooper and J. Graham (1995). "Active shape models - their training and application". Computer Vision and Image Understanding (61): 38–59.
4. Распознавание лиц на основе OpenCV для C++ (Facial Recognition based on OpenCV C++) // Проблема анализа изображений [электронный ресурс] //URL: [https://api-2d3d-cad.com/face\\_recognition\\_with\\_opencv/](https://api-2d3d-cad.com/face_recognition_with_opencv/) (дата обращения: 18.04.2021).
5. Georgia Tech Face Database [электронный ресурс] // URL: [http://www.anefian.com/research/face\\_reco.htm](http://www.anefian.com/research/face_reco.htm) (дата обращения: 18.04.2021).

С.А. Кудубаева<sup>1</sup>, Б.Т. Жусупова<sup>2</sup>

### СНЯТИЕ ОМОНИМИИ ПРИ КОМПЬЮТЕРНОМ ПЕРЕВОДЕ С КАЗАХСКОГО ЯЗЫКА НА ЯЗЫК ЖЕСТОВ НА ОСНОВЕ СЛОВАРЯ ЛЕКСИЧЕСКИХ ЗНАЧЕНИЙ МНОГОЗНАЧНЫХ СЛОВ

<sup>1</sup>Евразийский национальный университет имени Л.Н.Гумилева,  
<sup>2</sup>Костанайский региональный университет имени А.Байтурсынова)

В настоящей статье рассматривается метод снятия лексической неоднозначности в предложении на казахском языке при разработке компьютерной системы перевода с казахского языка на казахский жестовый язык (КЖЯ). При разработке данной системы перевода возникла необходимость учета семантической составляющей для получения качественного перевода, понятного носителям казахской жестовой речи.

Несмотря на то, что существуют корпуса казахского языка, достаточное количество компьютерных приложений, обрабатывающих казахский язык, не решена проблема, связанная с решением задач компьютерной лингвистики для казахского языка с необходимым уровнем качества обработки текста, снятием многозначности слов. Существование подобных проблем связано не только со сложностью обработки особенностей морфологии, синтаксиса и семантики казахского языка, но и с исторически сложившимся поздним началом работ по исследованию казахского языка.

Большая часть современных исследований в области автоматической обработки казахского языка направлена на автоматизацию его морфологического и синтаксического анализов [1–4], что во многом связано с реально существующими проблемами автоматической обработки агглютинативных языков.

Так как существующие корпуса казахского языка не содержат подкорпус со снятой омонимией и до этого не разрабатывались формализованные словари лексических значений многозначных слов, было принято решение исследовать методы снятия неоднозначности при обработке естественного языка с учетом особенностей казахского языка и составить формализованный словарь лексических значений омонимов и многозначных слов казахского языка, в некоторой степени аналогичный семантическому словарю В.Тузова [5]. Но так как словарь



В.Тузова отражает перевод на семантический язык слов русского языка и создание работающей компьютерной системы семантического анализа текстов на русском языке, имеющем флективный строи, противоположный агглютинативному, то алгоритм семантического анализа по словарю В.Тузова не может быть полностью перенесен на казахский язык. Возможно только использование некоторых методов формализации, чтобы построить аналогичную систему, как и предлагает сам автор.

Согласно общей схеме перевода с казахского языка на КЖЯ вводимое предложение проходит этапы морфологического и синтаксического анализов. Передача семантической составляющей языка перевода (казахского языка) происходит за счет снятия омонимии на этапе морфологического анализа с помощью формализованного словаря лексических значений омонимов и многозначных слов.

Методы обработки казахского языка определены, исходя из природы языка, относящегося к агглютинативным языкам, в которых словообразование происходит посредством агглютинации (приклеивания) к основе слова различных аффиксов (суффиксов и окончаний)

Пример:

**кітап – тар – ыңыз – дан – «из ваших книг»,**

для выражения значения множественного числа, значения притяжательности и падежных значений, используются три аффикса, следующих друг за другом.

При морфологическом анализе для каждого слова в предложении определяются его словообразующие и словоизменяющие морфемы и их характеристики, таким образом получаем набор морфологических характеристик по каждому слову в предложении. Основной алгоритм:

- 1) на вход подается предложение на казахском языке, текст считывается по-словно;
- 2) производим поиск совпадающих слов, при их нахождении, оставляем оба слова;
- 3) для каждого слова осуществляется поиск суффиксов и окончаний с конца слова, при этом используется приоритет групп аффиксов, характеризующих различные типы морфологических свойств (определяемый строгим порядком следования аффиксов в составе слова). Порядок убывания приоритета следующий: 1) падеж; 2) окончания формы принадлежности; 3) множественное число; 4) притяжательные окончания; 5) имя существительное; 5) имя прилагательное; 6) имя числительное; 7) наречие; 8) деепричастие; 9) глагол.
- 4) для каждого слова, в том числе и для совпадающих слов, определяем их принадлежность к омонимам и многозначным словам. Затем определяем альтернативы их использования согласно формализованному словарю лексических значений омонимов и многозначных слов.



5) В соответствии с найденным составом слова определяются часть речи и морфемный состав слова. Выводим лемму, часть речи и морфемы каждого слова.

В качестве программной среды для создания системы перевода с казахского языка на КЖЯ был выбран Visual Studio 2019, алгоритм и интерфейс написан на языке C#, имеющий большое количество методов для обработки текста.

База слов, суффиксов и окончаний хранится в базе данных MySQL, администрирование посредством phpMyAdmin, установленной на локальном сервере МАМР. В дальнейшем база данных будет подключена к хостингу сайта surdo.kz.

Формализованный словарь лексических значений омонимов и многозначных слов хранится в файле в формате UTF-8, так как остальные форматы не поддерживают казахский шрифт. Построение данного формализованного словаря омонимов основано на разработках лингвистов по казахскому языку: 1) 15-томного толкового словаря казахского литературного языка (разработанный исследователями Института языкознаний имени А.Байтурсынова), 2) словаря омонимов казахского языка «Қазақ тілінің омонимдер сөздігі», автора М.Белбаевой [6-7].

Именно посредством формального языка возможно описать смысл каждого омонима и многозначного слова, сделать данный словарь понятным компьютеру, удобным для обработки.

Фрагмент словаря представлен ниже:

АБАЙ {Жалқы Жанды} (Z1: !тәуелдік, Z1: !септік)

АБАЙ {Зат Жансыз} N%~НАЗАР (Z1: !ын)

АБАЙ {Үстеу} (Z1: !мен, Z2: !сыз,)

АБАЙ {Сын} N%~САҚ (Z1: !бол)

АБАЙ {Зат Жансыз} N%~БЕЛГІ (Z1: !бер)

В данном случае слово АБАЙ имеет 5 альтернатив: имя собственное одушевленное; существительное неодушевленное; наречие; прилагательное; существительное неодушевленное. В фигурных скобках указывается часть речи. В круглых скобках - условие для выбора данного значения омонима, таким образом, выбор альтернативы зависит от морфологических характеристик самого слова либо характеристик, находящегося рядом слова. Выбор верной альтернативы многозначного слова в дальнейшем в системе перевода заменяется на верный жест, тем самым передается правильное смысловое значение предложения для носителя жестовой речи.

### Литература

1. Gulshat Kessikbayeva, Ilyas Cicekli. Rule based morphological analyzer of Kazakh language / Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM. – 46-54 p.

2. В. Б. Барахнин, Л. Х. Лукпанова, А. А. Соловьев. Алгоритм построения словоформ с использованием флективных классов для систем морфологическо-



го анализа казахского языка / Вестник Новосибирского гос. университета. Серия: Информационные технологии, 2014. – Т.12.– № 2. – С. 25-32.

3. Тукеев У.А., Тургынова А. Морфологический анализ казахского языка на основе полной системы окончаний / Труды международной конференции по компьютерной и когнитивной лингвистике. Сер. "Интеллект. Язык. Компьютер" 2016. – С. 225- 231.

4. Мамырбаев О. Ж., Хайрова Н. Ф., Мухсина К. Ж. Қазақ тіліндегі мәтіндердегі қылмыстық мәнді коллакцияларды анықтау / Вестник КазАТК им. М. Тынышпаева, рекомендуемый ККСОН МОН РК. – №3(110). – 2019. – 170 -175 с.

5. Тузов В.А. Компьютерная семантика русского языка. — СПб.: Изд-во СПбГУ, 2004. — 400 с.

6. Қазақ әдеби тілінің сөздігі. Он бес томдық. / Құраст. Т.Жанұзақ, С.Омарбеков, Ә.Жүнісбек және т.б. – Алматы, 2011.

7. «Қазақ тілінің омонимдер сөздігі», М.Белбаева, 1988 – 193 с.

Д.С. Лебедева, А.И. Максимов

## ДИАГНОСТИКА ФАЛЬСИФИКАЦИИ ФОНОГРАММ НА ОСНОВЕ СИАМСКИХ СЕТЕЙ

(Самарский университет)

### Введение

Поскольку в настоящее время фонограммы активно используются как доказательство в судебных процессах, задача проверки аутентичности аудиозаписей в рамках криминалистической экспертизы является злободневной и актуальной. Под криминалистической экспертизой звукозаписи [1] понимается исследование магнитных или каких-либо других записей звука с целью установления фактов, имеющих доказательственное значение.

В данной работе исследуются возможности применения методов нейронных сетей и машинного обучения [2] для решения поставленной задачи при помощи сравнения фрагментов фоновых шумов аудиозаписей.

Фоновый шум [1] является частью общего шума, поступающего от подвижных или стационарно расположенных источников, при отключении известных источников.

В случае реальной диагностики цифровой фонограммы с использованием сравнения фрагментов фоновых шумов, эксперту необходимо выделить в анализируемой аудиозаписи фрагменты шума. В случае существенного различия метрик, характеристик, параметров взятых фрагментов шума, можно заключить, что в аудиозаписи присутствует вставка – ее части записаны в различных условиях. Для исследовательских целей упростим постановку задачи – пусть имеется пара фрагментов фонограмм фонового шума, необходимо установить, записаны ли эти фрагменты в одинаковых условиях.