УДК 004.82, 004.93

<https://doi.org/10.47533/2020.1606-146X.34>

С. А. *КУДУБАЕВА¹, Б. Т. ЖУСУПОВА², Д. Ж. АЛИППАЕВА²

*¹Евразийский национальный университет имени Л.Н.Гумилева,
г. Нур-Султан, Казахстан*

*²Костанайский региональный университет имени А.Байтурсынова,
г. Костанай, Казахстан*

saule.kudubayeva@gmail.com, botasha_zhan@mail.ru, alippaewi@mail.ru

СЛОВАРИ КАЗАХСКОГО ЯЗЫКА КАК ОСНОВА СЕМАНТИЧЕСКОГО АНАЛИЗА В СИСТЕМЕ КОМПЬЮТЕРНОГО СУРДОПЕРЕВОДА

Рассматривается вопрос разработки семантического словаря казахского языка для системы компьютерного перевода с казахского языка на казахский язык жестов, в которой будет учитываться семантика казахского языка и казахского жестового языка. В качестве основы технологии компьютерного перевода казахского языка на казахский жестовый язык служит семантический словарь казахского языка. В дальнейшем он позволит проводить семантический анализ исходного текста. Авторами статьи проведен анализ и подбор имеющихся словарей казахского языка, используемых при разработке базы данных семантического словаря. Словари казахского языка дают возможность для осуществления компьютерного сурдоперевода КЖЯ.

В статье также представлена возможность использования нотации Л.С.Димскис для разработки словаря структуры жестов казахского жестового языка. Раскрыта перспектива его включения в базу данных семантического словаря. А также раскрыта необходимость словаря жестов при разработке системы автоматизированного сурдоперевода в целом с учетом его эффективности и возможности полноценного практического использования.

Ключевые слова: *жестовый язык, семантический словарь, компьютерный сурдоперевод.*

Введение. Жестовый язык – это основное средство межличностной коммуникации большинства глухих и части слабослышащих людей [1]. Свои особенности есть в морфологии, синтаксисе, лексике жестового языка. Жестовые языки глухих разных стран отличаются своеобразием лексики и грамматического строя.

По своим коммуникативным функциям, не уступая звучащим языкам, жестовый язык является полноценным самостоятельным языком. Аналогично звучащим языкам, имеющим слова с различными значениями, жестовый язык содержит в себе однозначные, многозначные, разнозначные жесты. Полностью понять правильное значение жеста можно из контекста.

При разработке автоматизированной системы сурдоперевода казахского жестового языка (КЖЯ) очень важное значение при переводе имеет правильность передачи

*Адрес для переписки. E-mail: saule.kudubayeva@gmail.com

смыслового значения казахских слов, то есть его семантика. Необходимость создания системы компьютерного сурдоперевода обоснована тем, что использование субтитров не позволяет в полной мере решить вопрос понимания слабослышащими отражаемого на экране текста на казахском языке, так как он не отражает особенности жестового языка: грамматику, синтаксис, семантику.

Семантика (образовано от греческого слова: *σημαντικός* — обозначающий). Семантика – это раздел семиотики и лингвистики, в рамках которого изучаются знаки и знаковые системы как средства выражения значения и смысла. Простыми словами, семантика – наука, изучающая смысл слов.

Методология и результаты исследования. Исходные данные разрабатываемой системы сурдоперевода с учетом семантики представляют собой текстовую строку.

Современные системы компьютерного перевода часто основываются на трехчленных моделях, одним из примеров которых может служить модель, предложенная З.М. Шаляпиной.

Общая схема компьютерного перевода казахской речи (текста) на КЖЯ основывается на анализе казахского текста и включает следующие этапы:

1. Морфологического анализа слов в предложении.
2. Синтаксического анализа структуры предложения и первичного семантического анализа.
3. Семантического анализа слов в предложении [2].

Основой технологии компьютерного перевода казахского языка на КЖЯ служит семантический словарь казахского языка. В дальнейшем он позволит проводить семантический анализ исходного текста.

Семантический словарь казахского языка содержит в себе несколько словарей: грамматический словарь, словарь лексических значений казахского языка, словарь фразеологизмов, словарь многозначных слов и омонимов, словарь синонимов, словарь предлогов и другие. Так как семантический словарь для компьютерного сурдоперевода представляет собой базу данных взаимосвязанных таблиц «Словарные статьи», «Семантические описания», «Слова», «Фразеологизмы», «Омонимы», «Синонимы», «Предлоги» и др., то в ней необходимо также наличие таблицы «Жесты» и таблицы-связки «Жест - слово» для дальнейшего определения соответствия «слово-жест».

За основу словаря лексических значений, входящего в семантический словарь казахского языка взят 15-томный толковый словарь казахского литературного языка, содержащий 92 300 слов и 57 856 словосочетаний [3]. Фрагмент данного словаря представлен в Таблице 1.

Таблица 1 – Фрагмент словаря лексических значений казахского языка

Слово	Лексическое значение (15-томный словарь казахского литературного языка)	Словосочетания
1	2	3
Адам / Денсаулық, медицина		
Ана	зат. 1. Баланың туған шешесі; балалы әйел. 2. Елге сыйлы, жасы үлкен әйел адам.	Ана сүті аузынан кетпеген. Ана сүті. Ана тілі.

Окончание таблицы 1

1	2	3
	3. Ауыс.поэт. Негіз, түп, тамыр. 4. Есім.Со-нау, ана бір, анау.	Ана махаббаты. Ана сүтін ақтады.
Апа	Зат. 1. Баланың туған анасы, шешесі. 2. Бірге туған қыздардың үлкені, әпке. 3. Жерг. әкесінің шешесі, әжесі. 4. Жерг. жасы үлкен әйел адамды құрметтеу мағынасында айтылатын қаратпа атауыш.	Нағашы апа.
...	...	...
Адам / Киім, аяқ киім, сән		
Аяқ киім	Зат. Аяққа киетін киімдердің жалпы атауы	Аяқ киімін киді. Аяқ киім сатып алды.
Белдемше	Зат. 1. <француз. Жүре> той-думанда көйлек үстінен киетін әйел киімінің сәнді түрі. 2. Тазалық үшін алға ұстайтын алжапқыш. Зат. 2. Ортан бел, белуар.	Белдемше кию, белдемшесін байлады
...	...	...

Словарь синонимов, содержащий 8246 слов и доступный на сайте lugat.kz, и словарь «Синонимдер сөздігі» автора Бизакова С. (Алматы, 2007) являются основой словаря синонимов казахского языка, являющего частью разрабатываемого семантического словаря [4], фрагмент его представлен в Таблице 2.

Таблица 2 – Фрагмент словаря синонимов казахского языка

Слово	Синонимы
Ауру	науқас, кесел, дерт, сырқат
Бас	Әуелі. Алғашқы. Басшы, жетекші.
Дене	Адам, дене тұрқы
Емхана	Аурухана
Қолғап	Биялай
Орамал	Жаулық, сүлгі, шүберек
Сән	Мода
Сөмке	Дорба
Шалбар	Сым (диалект)
Ана	Апа, шеше
Ата	Қария, карт, ақсақал.
Бала	Нәресте, сәби, бөбек.
...	...

Морфологический модуль системы сурдоперевода должен содержать следующие морфологические словари: грамматический словарь, словарь имен собственных, сло-

варь географических мест. Для компьютерной обработки казахского языка важно, прежде всего, формирование морфологического словаря, еще не созданного на казахском языке на уровне, требуемом системой обработки текста.

Одна из коммерческих IT-фирм, которая начала работать в стране по компьютеризации казахского языка, охарактеризовала морфологический словарь казахского языка следующим образом: для дальнейшего синтаксического анализа в ресурсах системы создан морфологический словарь казахского языка, который используется для правильного системного определения нормального состояния слов в тексте, грамматических атрибутов. В морфологический словарь казахского языка включены необходимые части слов, а также по каждому слову включены все формы слов, которые возникли с учетом закона сингармонизма.

Основой создания морфологического словаря казахского языка служат следующие труды: «Қазақ тілінің функционалды грамматикасы» (Алматы, 2012), «Қазіргі қазақ тілінің морфологиясы» (Алматы, 2007) автора Оралбай Н. и «Қазіргі қазақ тілінің морфемалар жүйесі» (Алматы, 2001) авторов Оралбаева Н., Қалыбаева А.

Для формирования словаря имен собственных и словаря географических мест используются словари, разработанные Институтом языкознания имени А.Байтурсынова:

- Жанұзақов Т. «Қазақ тіліндегі жалқы есімдер» (Алматы, 1965);
- Жанұзақов Т., Есбаева К. «Қазақ есімдері» (Алма-Ата, 1988);
- Т.Жанузаков, К.К.Рысбергенова, Н.Б.Онгарбаева «Қазақстан Республикасының топонимдері» (Алматы, 2001);
- «Қазақ жер-су аттары» Энциклопедиялық анықтағыш (Алматы, 2009);
- «Қазақ кісі аттары» Энциклопедиялық анықтағыш (Алматы, 2009).

На их основе создаются соответствующие таблицы в базе данных семантического словаря.

Результат морфологического анализа является входной информацией для синтаксического и первичного семантического анализа, в ходе которых определяются семантические отношения.

Для разрешения проблемы лексической многозначности необходимо обработать омонимы и фразеологизмы в предложении. Омонимия – это совпадение по звучанию и написанию различных слов. Фразеологизмы отличаются от обычных сочетаний слов тем, что общее значение фразеологического оборота не равно сумме отдельных значений слов. В семантическом словаре казахского языка словари многозначных слов или омонимов и фразеологизмов основываются на следующих словарях:

- «Қазақ тілінің омонимдер сөздігі» М.Белбаева (1988);
- «Мағыналас фразеологизмдер сөздігі» Г. Смағұлова (Алматы, 2010);
- «Қазақ тілінің фразеологиялық сөздігі» І.Кенесбаев (Алматы, 2007).

Разработка базы данных семантического словаря, включающего в себя взаимосвязанные таблицы перечисленных словарей казахского языка, является основной задачей при создании семантического модуля системы компьютерного сурдоперевода. Таким образом, словари казахского языка являются основой, дают возможность для осуществления компьютерного сурдоперевода КЖЯ.

Учитывая то, что при разработке системы компьютерного сурдоперевода жестового языка необходимо переводить текст на язык жестов, распознавать жесты, для

этого требуется наличие в базе данных семантического словаря таблицы по жестам КЖЯ. Словарь жестового языка - основа любого компьютерного сурдоперевода.

Однако следует отметить, что не существует такого словаря КЖЯ, который позволял бы по форме жеста находить его значение, хотя в мире создание подобных словарей жестовых языков – достаточно распространенная практика. Учитывая возможность применимости системы Л.С. Димскис к КЖЯ, нами было проверено достаточно ли компонентов (знаков) жестовой нотации Л.С.Димскис для точного описания жестов казахского ЖЯ, необходимы ли дополнительные знаки для записи жестов [5].

Разработанный словарь структуры жестов КЖЯ может быть использован для осуществления поиска слов в словаре, представления их в виде нотации и отображения с помощью аватара. Также словарь жестов позволит осуществлять распознавание КЖЯ. При этом подходе, представляя жесты в виде компьютерных нотаций, проводя поиск слов в словаре, лингвистическую обработку, результат распознавания представляется в виде текста или голосового сообщения.

Проводимые исследования по анализу и подбору словарей казахского языка, для создания семантического словаря системы компьютерного сурдоперевода казахского жестового языка обоснованы необходимостью обеспечить комфортную и доступную среду для безбарьерного общения людей с нарушением слуха со слышащими, в том числе при получении образования, медицинских и государственных услуг и прочее.

ЛИТЕРАТУРА

1 Зайцева Г.Л. Жестовая речь. Дактилология: Учебник для студентов высших учебных заведений / Г.Л. Зайцева. – М.: Владос, 2000. – 192 с. [Zajceva G.L. Zhestovaya rech'. Daktilologiya: Uchebnik dlya studentov vysshih uchebnyh zavedenij / G.L. Zajceva. – М.: Vlados, 2000. – 192 s.]

2 Мануева, Ю. С., Гриф, М. Г., & Козлов, А. Н. (2014). Построение системы компьютерного сурдоперевода русского языка. Труды СПИРАН, 6(37), 170-187. <https://doi.org/10.15622/sp.37.11> [Manueva, YU. S., Grif, M. G., & Kozlov, A. N. (2014). Postroenie sistemy komp'yuternogo surdoperevoda russkogo yazyka. Trudy SPIRAN, 6(37), 170-187. <https://doi.org/10.15622/sp.37.11>]

3 Қазақ әдеби тілінің сөздігі. Он бес томдық. / Құраст. Т.Жанұзақ, С.Омарбеков, Ә.Жүнісбек және т.б. – Алматы, 2011. [Qazaq adebi tilining sozdigi. On bes tomdyq. / Qurast. T.Zhanuzaq, S.Omarbekov, Ә.Zhunisebek zhane t.b. – Almaty, 2011.]

4 Бизақов С. Синонимдер сөздігі. Алматы: «Арыс» баспасы, 2007. – 640 б. 5 Димскис, Л. С. Мы изучаем жестовый язык. Москва: Академия. – 2002. – 128 с. [Bizaqov S. Sinonimder sozdigi. Almaty: «Arys» baspasy, 2007. – 640 b. 5 Dimskis, L. S. My izuchaem zhestovyy yazyk. Moskva: Akademiya. – 2002. – 128 s.]

С. А. КУДУБАЕВА¹, Б. Т. ЖУСУПОВА², Д.Ж. АЛИППАЕВА²

*¹Л.Н.Гумилев атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан қаласы, Қазақстан*

*²А.Байтұрсынов атындағы Қостанай өңірлік университеті,
Қостанай қаласы, Қазақстан*

**КОМПЬЮТЕРЛІК СУРДОАУДАРМА ЖҮЙЕСІНДЕГІ СЕМАНТИКАЛЫҚ
ТАЛДАУДЫҢ НЕГІЗІ РЕТІНДЕ ҚАЗАҚ ТІЛІНІҢ СӨЗДІКТЕРІ**

Бұл мақалада қазақ тілі мен қазақ ымдау тілінің семантикасы ескерілетін қазақ тілінен қазақ ымдау тіліне компьютерлік аударма жүйесі үшін қазақ тілінің семантикалық сөздігін әзірлеу мәселесі қарастырылады. Қазақ тілінен қазақ ымдау тіліне компьютерлік аударма технология негізі ретінде қазақ тілінің семантикалық сөздігі қызмет етеді. Болашақта ол бастапқы мәтінге семантикалық талдау жүргізуге мүмкіндік береді. Мақала авторлары семантикалық сөздіктің деректер базасын әзірлеуде қолданылатын қазақ тілінің сөздіктеріне талдау және іріктеу жүргізді. Қазақ тілінің сөздіктері қазақ ымдау тілінің компьютерлік сурдоаудармасын жүзеге асыруға мүмкіндік береді.

Мақалада сондай-ақ қазақ ымдау тілінің ымдау құрылымының сөздігін әзірлеу үшін Л.С. Димскистің нотациясын пайдалану мүмкіндігі ұсынылған. Оны семантикалық сөздіктің деректер базасына енгізу перспективасы ашылды. Сондай-ақ автоматтандырылған сурдоаударма жүйесін әзірлеу кезінде оның тиімділігі мен толыққанды практикалық пайдалану мүмкіндігін ескере отырып, қимыл сөздігінің қажеттілігі ашылды.

Түйін сөздер: ым тілі, семантикалық сөздік, компьютерлік сурдоаударма.

S. A. KUDUBAYEVA¹, B. T. ZHUSSUPOVA², D. ZH. ALIPPAYEVA²

¹ L.N. Gumilyov Eurasian national University, Nur-Sultan, Kazakhstan

² Kostanay regional university named after A. Baitursynov, Kostanay, Kazakhstan

DICTIONARIES OF THE KAZAKH LANGUAGE AS THE BASIS OF SEMANTIC ANALYSIS IN THE SYSTEM OF COMPUTER SIGN LANGUAGE TRANSLATION

This article discusses the development of a semantic dictionary of the Kazakh language for a computer translation system from Kazakh to Kazakh sign language, which will take into account the semantics of the Kazakh language and the Kazakh sign language. The semantic dictionary of the Kazakh language serves as the basis of computer translation technology from the Kazakh language to the Kazakh sign language. In the future, it will allow semantic analysis of the source text. The authors of the article analyzed and selected the available dictionaries of the Kazakh language used in the development of the semantic dictionary database. Dictionaries of the Kazakh language provide an opportunity for computer-based sign language translation of the Kazakh sign language.

The article also presents the possibility of using L. S. Dimskis notation to develop a dictionary of the structure of gestures of the Kazakh sign language. The prospect of its inclusion in the database of semantic dictionary is revealed. And also revealed the need for a dictionary of gestures in the development of automated sign language translation system as a whole, taking into account its effectiveness and the possibility of full practical use.

Key words: sign language, semantic dictionary, computer sign language translation.