

Д. Рахимова А., Тұрарбек, В. Карюкин, А. Карибаева, А. Тұрғанбаева

**ҚАЗАҚ ТІЛІНЕ АРНАЛҒАН ЗАМАНАУИ МАШИНАЛЫҚ АУДАРМА
ТЕХНОЛОГИЯЛАРЫНА ШОЛУ**

I. Кіріспе

«Қазақтың тілі – қазақтың жаны! Халқымыздың тарихы да, тағдыры да – тілінде. Мәдениеті мен әдебиеті де, ділі мен діні де – тілінде», деп президентіміз Қ.Тоқаев айтып, қазақ тілінің мемлекетіміздің дамып, өркендеуіне, әлемдік дәрежеде танылуына қатысы бар екенін атап өткен.

Қазіргі әлем және біздің жақын болашағымыз машиналық оқытуға байланысты, өйткені күн сайын жаңа технологиялар көбеюде. Машиналық оқытудың міндеттерінің бірі – бір табиғи тілден екінші тілге аудару. Машиналық аударма адамдарға тілдердің әртүрлігіне қарамастан сөйлесуге мүмкіндік береді, өйткені бұл тілдік кедергіні бұзып, жаңа қарым-қатынас тілдерін ашады. Машиналық аударма – бұл техниканың жаңалығы, адамзат дамуындағы ерекше қадам. Аударманың бұл түрі сіздің сұхбаттасушының хатта не жазғанын немесе не айтқанын дереу түсіну қажет болған кезде көмектесе алады. Әрине, мұндай аударманың сапасы өте төмен, бірақ көп жағдайда басты мағынаны түсінуге болады. Толық мағынасы қажет болғанда не істеуге болады? Мұнда қолмен аударма көмекке келеді.

Әлемде әр түрлі қызмет салаларында қолданылатын арнайы тілдер бар, сол тілдердің көмегімен іскерлік келіссөздер жүргізіліп, ғылыми зерттеулер және тағы басқа әрекеттер жасалып жатыр. Қазір әлемде 6000-нан астам тіл бар, олардың үштен бірі дамып келе жатқан аймақтардың тілдері болып табылады, бұл жерде тілдік тосқауыл сыртқы нарыққа шығу үшін күрделі мәселе болып табылады. Машиналық аударма технологиясының заманауи дамуы қолданушыларға осы мәселені шешуге бір қадам жақындауға мүмкіндік берді. Машиналық аударманы пайдаланушы компанияның кез-келген қызметкері, инженер, заңгер, дәрігер, мұғалім, менеджер және басқалар бола алады, өйткені ол көп парақтық құжаттардың сапалы аудармасын қысқа мерзімде ала алады.

Біздің елде машиналық аударма 2009 жылдан бастап дами бастады. Қазақ тіліне машиналық аударманы алғашқылардың бірі болып У.А.Төкеев пен А.А.Шәріпбаев бастады. У.А.Төкеевтің машиналық аударма саласында ғылыми зерттеулермен белсенді айналысатын ғылыми мектеп құра алды. Отандық шәкірттерінің ғылыми зерттеулерінде Д.Р.Рахимованың жасаған орыс тілінен қазақ тіліне машиналық аударма алгоритмін, А.Ж.Картбаевтың жасаған морфологиялық сегментация әдісін қолдана отырып, ағылшынша-қазақша сөздерді туралаудың статистикалық моделін атап өтуге болады.

У.А.Төкеевтің ғылыми тобы 2009 жылдан бастап «Разработка математической модели и программного продукта компьютерного перевода казахского языка на английский (простые предложения)» жобасы нәтижесінде қазақ тілінен ағылшын тіліне жай сөйлемдер грамматикасының формальді модельдері және қазақ тілінен ағылшын тіліне машиналық аударма программасының алғашқы нұсқасы құрастырылды [1-2]. 2012-2014 жылдары «Разработка эффективных технологии компьютерного перевода казахского языка на английский и русский языки (и обратно) на основе методов формальных грамматик и статистических методов» жобасы нәтижесінде орыс, қазақ тілі сияқты морфологиялық күрделі табиғи тілдердің машиналық аударма үшін көпмағыналы бейнелеу әдісі құрастырылды [3-5]. 2015-2017 жылдары Аликанте Университетінің (Испания) профессоры М.Форкаданың ғылыми тобымен бірлесе отырып «Разработка свободной/открытой системы машинного перевода с казахского языка на английский и русский языки (и обратно) на базе платформы Apertium» жобасы нәтижесінде Apertium платформасының базасында қазақ тілінен ағылшын тіліне және орыс тіліне (және керісінше) грамматикалық ережелерге негізделген ашық аударма жүйесі жасалынды [6-9]. 2018 жылдан бастап «Разработка и исследование системы нейронного машинного перевода казахского языка» жобасы бойынша жұмыстар жүргізіліп жатыр. Дегенмен соңғы үш-төрт жылдағы машина аудармасының теориясы мен тәжірибесі едәуір өсті, машина аудармасының өте жоғары сапаны көрсететін жаңа бағыты жасалды. Бұл сәйкесінше өнеркәсіптік машиналық аудару сапасының стандартын жаңа биіктікке көтерді [10-13].

Мақаланың мақсаты – машиналық аударма технологияларына шолу жасау, осы технологияларды қазақ тілінде қолдану және аудармашы қосымшаларда машиналық аударма қателіктерін анықтау.

II. Әдістер.

Машиналық аударма дегеніміз – мәтінді (жазбаша және ауызша) арнайы компьютерлік бағдарламаның көмегімен бір табиғи тілден екінші тілге аудару үрдісі [14]. Алғашқы ЭЕМ пайда болуымен бірге мәтіндерді автоматты түрде аудару туралы айтыла бастады. Себебі, машиналық аударманы үлкен деректерді аудару қажет болған кезде және көптеген бағыттарда қолданады, оны автоматтандыру бұл күнделікті өмірде ғана емес, көптеген салалар үшін пайдалы болар еді.

Қазіргі уақытта машиналық аударманың үш технологиясы бар:

– аналитикалық машиналық аударма;

– статистикалық машиналық аударма;

– нейрондық машиналық аударма.

Аналитикалық машиналық аударма (RBMT, rule-based machine translation, ережеге негізделген машиналық аударма) [15] – машиналық аударманың алғашқы технологиясы. Бұл әдіс екі тілдегі аударма ережелерінің жиынтығын және лексикалық элементтер жиынтығын пайдаланады. Осы әдіс бойынша талдау келесі кезеңдерден өтеді:

1. Морфологиялық талдау ол сөздердің түрі, тегі және басқа морфологиялық сипаттамалары бойынша анықталады;

2. Жеке сөздерді топтарға біріктіру;

3. Сөйлемдердің синтаксистік талдауы, яғни сөйлем мүшелерін және олардың сөйлемдегі орнын анықтау;

4. Сөйлемдер синтезі – сөйлемдердің табылған бөліктерін және тәуелді сөздер тобын үйлестіру.

Бұл әдістің кемшіліктері артықшылықтарынан көп болып келеді. Өйткені, бұл процесс көп еңбек пен уақытты қажет етіп, лингвистикалық деректер қорын үнемі жаңартып отыруды және тіл мамандарын тартуды керек етеді. Сонымен бірге, аналитикалық аударма синтаксистік және морфологиялық дәлдікті, нәтиженің тұрақтылығын, пәндік аймаққа бейімделуді қамтамасыз етеді.

Статистикалық машиналық аударма (corpus-based machine translation, CBMT) – мәтіндік корпуста негізделген машиналық аударма әдісі [15]. Әдістің негізі дайын ережелермен жұмыс істемейді, ол параллельді екі тілді мәтіндік корпустарға негізделген ережелерді құру болып табылады.

Байес теоремасына негізделген бұл технологияда аударманың барлық нұсқалары қарастырылады және олардың әрқайсысының ықтималдығы ескеріледі. Статистикалық аударма жүйесі әдетте үш компоненттен тұрады:

1. Аударма моделі дегеніміз – жүйеде барлық белгілі сөздер мен сөз тіркестерінің бір тілде болуы мүмкін аудармалар тізімін беретін және осы аудармалардың ықтималдығын көрсететін сөздік кесте;

2. Тілдің ықтималды моделі дегеніміз – мәтін аударылатын тіл туралы жүйенің білім көрінісі;

3. Декодер – тікелей аудармамен айналысатын аудармашының құрамдас бөлігі. Бастапқы мәтіннің әр сөйлемі үшін ол аударманың барлық нұсқаларын тандайды, аударма моделінен сөз тіркестерді біріктіреді және оларды ықтималдылықтың азаю ретімен сұрыптайды.

Нейрондық аударма (нейрондық машиналық аударма, NMT) – бұл ең жаңа «тәсіл», ол басқа технологияларға қарағанда жоғары сапалы аударма береді. Технология жасанды нейрондық желілерді [16], дәлірек айтқанда, екі бағытты қайталанатын нейрондық желілерді қолдана отырып, белсенді аударма әдістеріне негізделеді. Нейрондық желінің негізгі мәселесі – бұл сапалы жүйені аудару үшін көп жағдайларды, кем дегенде 500 миллион токендерді пайдалану қажеттілігі және осындай жүйені оқыту үшін үлкен есептеу қуатын талап етуі. 2015-2016 жж. аралығында осындай жүйелерді құрастырушылардың барлығы нейрондық аудармашылардың нұсқаларын бірінен соң бірі ұсына бастады. Сонымен, 2017 жылы Google және Яндекс нейрондық аудармаға біртіндеп көшу туралы жария етті. Сондай-ақ Linex сервер платформасында ішкі нейрондық аудармасы бар PROMT өнімдері және Windows Server үшін нейрондық аудармасы бар PROMT аударма сервері бар. Нейрондық желілердің артықшылықтары мен кемшіліктеріне келетін болсақ, олар көбінесе статистикалық жүйелерге ұқсас. Машиналық аударманың негізгі құралы ретінде нейрондық желілердің айқын болашағына қарамастан, қазіргі уақытта әртүрлі тақырыптардағы мәліметтердің тиісті мөлшерінің болмауы машиналық аударманы қолданудың барлық саласында нейрондық желілердің сөзсіз артықшылығы туралы айтуға мүмкіндік бермейді [17].

III. Нәтижелер.

Қазақ тілі – күрделі номинативті (морфологиялық және синтаксистік), полисинтетизмнің қатысуы бар агглютинативті жұрнақты тіл. Мемлекетіміздің жылдан жылға әлемдік деңгейде дамып, шетелдік байланыстардың артуына байланысты қазақ тіліне немесе қазақ тілінен басқа тілдерге аударуда әртүрлі аудармашы бағдарламалар кеңінен қолданылып келеді. Аудармашылар толық дұрыс аударуы әлі де мүмкін емес, аудару кезінде қателіктер болып отырады, дегенмен машиналық аударма саласы алдыңғы жылдардан қарағанда әлде қайда дамыған.

Машиналық аударма қателіктерін талдау үшін ең танымал онлайн аудармашылар таңдалды. Орыс тілінен қазақ тіліне аударғанда әр түрлі қателіктер пайда болуы мүмкін, өйткені қазақ тілі басқа тілдерден ерекшеленеді, ол ерекше сипаттамаларға ие:

– лексикалық құрамның жақындығы;

– сингармонизм заңы;

- агглютинация - аффикстердің бірізді жалғануы;
- санаттың болмауы;
- көмекші сөздер (предлогтардың) болмауы;
- арнайы сөз тәртібі.

Машиналық аударма кезінде туындайтын қателіктерді анықтау үшін ең алдымен аудармашы бағдарламаларда қарапайым хабарлы сөйлемдерді аударып көрейік. Кесте 1 қазақ тілінен орыс тіліне аударылатын сөйлемдер мен аударма нәтижелері көрсетілген. Жалпы жасалған талдауда 8 аудармашы бағдарлама қарастырылған. Кесте 1 алғашқы 4 түрін қамтиды.

1-кесте. Қазақ тілінен орыс тіліне машиналық аударма нәтижелері

Аудармашылар	Google	Яндекс.Аудармашы	Tridentsoftware	prof-translate.ru
Мектепке мен ертен барамын Анам мен әпкем жолда тұр Жолда тұрған анам мен әпкем	Я иду в школу завтра Моя мама и сестра уже в пути Мои мама и сестра в пути	И завтра иду в школу Мама и сестра стоят на дороге Мама и сестра, стоявшая на дороге	И завтра иду в школу Мама и сестра стоят на дороге Мама и сестра, стоящие в строке	И завтра иду в школу И завтра с утра иду в школу Мама и сестра стоят на дороге Мама и сестра, стоящие в строке
Мектепке мен ертен барамын Мен мектепке ертен барамын Мен кеше мектепке бармадым Ол кеше мектепке бармады Ол қазір мектепте	Я иду в школу завтра Я иду в школу завтра Я вчера не ходила в школу Он вчера не ходил в школу Она сейчас в школе	Я иду в школу Я иду завтра в школу Я не ходил в школу Этот человек не пошел в школу О қазір мектепте Он сейчас в школе	И завтра иду в школу И завтра иду в школу И вчера я не ходил в школу Он вчера в школу не ходил Он сейчас в школе	И завтра иду в школу И завтра иду в школу И вчера я не ходил в школу Он вчера в школу не ходил Он сейчас в школе
Жолдан қасқырды көрдік Жолда ешкім болған жоқ Бұл айналым жолы Басқа жол болмады	Мы видели волка на дороге На дороге никого не было Это объезд Другого пути не было	С дороги видели волка Никто не остался на дороге Путь этот оборот Не было другого пути	Строки видели волка На дороге никого не было Путь этот оборот Другого пути не было	Строки видели волка На дороге никого не было Путь этот оборот Другого пути не было
Алматы Көктөбе Қарағанды Үштөбе Жадыра Айкүміс	Алматы Коктобе Караганда Не беспокойся Жадыра Серебряный	Г. Алматы Коктобе Г. Караганда Уштобе Жадыра Айкумис	Алматы Коктобе Караганда Уштобе Жадыра Айкумис	Алматы Коктобе Караганда Уштобе Жадыра Айкумис
Желіде кемшіліктер бар болған Артышылықтардың тізімі көрсетілген еді Биіктен аққан ақ сәуле Арналарымыздың түрілері шектеулі Қанаттарымыздың ұзындығы анықталмаған	Были проблемы с сетью Был предоставлен список преимуществ Белый свет течет сверху Типы наших каналов ограничены Длина наших крыльев неизвестна	Есть недостатки в сети Был показан список преимуществ Ао утечки излучения с высоты Ограниченные типы каналов Длина крыльев не определена	Недостатки есть в сети Список преимуществ было указано в Ао утечки излучения с высоты С Арналарымыздың түрілері Длина Қанаттарымыздың не установлена	Недостатки есть в сети Список указанных Артышылықтардың бы Ао утечки излучения с высоты С Арналарымыздың түрілері Длина Қанаттарымыздың не установлена

Енді қателіктерге сүйене отырып, қазақ тілінен орыс тіліне аударғанда аудармашылар қандай жағдайларда қателескенін анықтауға тырысамыз.

1) «Мен» дегенді білдіретін «Я» сөзінің қазақ тіліне аудармасы. Кейбір жағдайларда аудармашы оны «и» көмекші сөзі ретінде аударған. Әр түрлі нұсқада аударуға тырысып, оны сөз тіркестерде қолданамыз. Нәтижесінде олар аударма кезінде уақыт сөзінің тіркесімімен қате жіберілді деген тұжырымға келді. Мысалы, аудармашылар (2-7) *Мен бүгін бармадым, ал И сегодня бармадым*. Қалған аудармашылар аударманың басқа нұсқаларын аударма білді. Бұл жерде «бармадым» сөзіне назар аударамыз. Барлық аудармашылар *Мен бүгін* сөз тіркесін дұрыс аударды.

2) Полисемантикалық сөздерді аудару. «Жол» сөзінің қазақ тіліндегі аудармасы көп мағыналы сөз болып табылады және «Путь», «Дорога», «Строка» деп аударылады. Аударманы дұрыс орындалмады (1,5,6,7,8). Аударманы тексеру үшін «Жол» сөзін жаңа тіркестермен толтыруға тырысамыз. Аударманың әртүрлі нұсқаларында «Путь» пен «Дорога» болып аударылды. Ең жақсы аударма (1,7,8). Көпше түрде аударуда көптеген қателіктер кездеседі.

3) Аудармадағы тегінің анықталмауы. Қазақ тілі сөйлемдерінде тегі (род) болмағандықтан, аудармашылар ол бойынша көбінесе әртүрлі қателіктер жібереді. «Ходить» сөзін қолдана отырып, тексеру жүргізейік. *Ол бармады - Она не ходила*. Барлық аудармашылар ер текті аударылған және барлығы (2,3,4,5,6,7) келесі түрде аударылған, яғни «в» көмекші сөзі қосылған. Текті анықтау үшін «күйеуіммен» сөзін қосамыз. Нәтижесінде (1)-ден басқа барлық аудармашылар дұрыс аудармады, *Он не ходил в күйеуіммен*, дұрыс аударып, текті анықтағанда, *Она не пошла с мужем* болды.

4) Тиісті есімдерді аудару. Аудармашы қателері анықталды (1,5,8). *Айкүміс* есімі *Серебряный* деп аударылған. Ұқсас есімдер аударылды. *Үштөбе* қаласының атауы *Не беспокойся* деп аударылған. Қазақстандық аудармашылардың барлығы есімдерді дұрыс аударды.

5) Көпше түрдегі аударма. *Арналарымыздың және Қанаттарымыздың* сөздері қате аударылған. Аудармашылар аударма алмады (3,4,5,6,7). Бұл сөздердің көпше түріне байланысты екенін анықтау үшін, бұл сөздерді басқа жағдайларда аударуға тырысамыз. Нәтижесінде, көптеген аудармашыларда көптік жалғау қосқанда, аудармашылар аударма алмайтыны анықталды.

6) Әдеби мәтінді аударудағы қателер. *Биіктен аққан ақ сәуле* сөйлемін аудару кезінде тек аудармашылар (1,8) дұрыс аударған, қалған аудармалар *Ао утечки излучения с высоты* түрінде аударды. «Ақ» сөзі қате аударылды, қатені анықтау үшін оны әртүрлі сөйлемдер мен тіркестерде аударуға тырыстық.

Қазақ тілінен орыс тіліне және орыс тілінен қазақ тіліне аударғанда әртүрлі қателіктерді болдырмас үшін сөйлемдердің морфологиялық талдауына жүгіну керек. Морфологиялық талдау – бұл табиғи тілге байланысты әртүрлі тапсырмалардың бастапқы кезеңі, сондықтан оның нақты орындалуы үлкен мәнге ие. Морфологиялық талдау әдістерін 3 түрге бөлуге болады:

- аффикстер сөздігімен талдау;
- аффикстер мен негіздердің сөздігін пайдаланып талдау;
- сөздер жүйесінің сөздігін пайдаланып талдау.

Енді бір сөйлемнің морфологиялық талдауын жасайық. *Жолдан қасқырды көрдік*.

1. *Жолдан* – қайдан? Пысықтауыш, шығыс септік

2. *Қасқырды* – кімді? Толықтауыш, табыс септік

3. *Көрдік* – не істедік? Баяндауыш, түбір – көр, етістік, ді – жұрнақ, к – жіктік жағау, 1 жақ, көпше түр. Мұнда *біз* жасырын тұрған бастауыш, жіктеу есімдігі.

Аудару нәтижесінде *С дороги видели волка және Строки видели волка* деген аударма алдық, мұнда аудармашылардың қандай келіктер жасағаны көрініп тұр. Орысша-қазақша аударма үшін қолданылатын сөйлемдер мен аудармашылар бірдей тәртіпте қалды. Нәтижесінде аудармашыларда қателіктерді жіктеп анықтадық:

1) Сөздердің көпмағыналығы. Туыстық қатынастарды аударудағы қателер. «Сестра» сөзі қазақ тіліне «Әпке» немесе «Апа», ал «Сестренка» сөзі «Қарындас» деп аударылған. Бірақ орыс тілінде «кіші қарындас» деген сөз жиі қолданылады, қазақша «Қарындас» немесе «Сіңлі» болады. Аударманы тексергеннен кейін аудармашылардың барлығы қателіктермен аударатыны байқалды, «Тесть» сөзін аудармашының 1 ғана дұрыс аударғаны анықталды.

2) Уақыт-мезгілге тәуелді сөздерді аудару. Мысалда «Тұру» сөзі аударылмаған, дұрыс нұсқасы - «Турсаң». Салыстыру үшін «Стендтер» және «Шыбындар» сөздерін алып, оларды уақытында өзгертіңіз. Аудармашылар дұрыс аударылған (1,2,8).

3) Есімдіктері бар сөздер. «Не ходил», «Бармадым» немесе «Бармады».

4) Тиісті есімдер. Айкумис есімінің аудармасы, аудармашыларда өзгеріссіз қалды (1,2,5,7).

5) Басқа тілдерден алынған сөздер. «Проблема» сөзін аудару кезінде тек 1 аудармашы ғана аударды, қалғандары сөзді өзгеріссіз қалдырды.

6) Сөйлемдерді жазу тәртібін сақтамау. Бұл жағдайда сөздер дұрыс аударылған, бірақ дұрыс орналастырылмаған, нәтижесінде жалпы мағынасы жоғалады.

Аудармашылардың жұмысын тексерудегі келесі қадамында үлкен көлемді ғылыми және публицистикалық мәтіндерді аудару болды, оны Кесте 2 көруге болады. Барлық 8 аудармашыға қазақ тілінен орыс және ағылшын тілдеріне, сондай-ақ орыс және ағылшын тілдерінен қазақ тіліне аудармалар жасалды.

2-кесте. **Яндекс аудармашы**

Жанр	Мәтін	қаз. – орыс	қаз. – ағыл.
Ғылыми	Бүгінгі Алматы қаласының геологиялық кезең бойынша ертедегі палеозой дәуірен бергі үштік кезең яғни мезозой дәуірінен бергі 60 000 000 жыл бұрын планетадағы Тетис теңізінің түбі болғаны геологиялық мәліметтерден белгілі. Оған дәлел Памир тауларының биік шыңдарында балықтардың сүйек қаңқалары табылып жатқандығы бір дәлел болса, екінші дәлел өзіміздің кәдімгі Каратаудың жоталарында әлі зерттелмей жатқан түрлі моллюскалардың миллиондаған қалдықтары дәлел іспеттес.	По геологическим данным, на сегодняшний день на планете было дно Тетисского моря-60 000 000 лет назад, начиная с древней палеозойской эпохи. Об этом свидетельствует тот факт, что на возвышенных вершинах Памирских гор находят костные скелеты рыб, второй доказательством тому является миллионы останков различных моллюсков, еще не исследуемых на хребтах нашего обычного Каратау.	According to geological data, today the planet had the bottom of the Tethys sea-60,000,000 years ago, starting from the ancient Paleozoic era. This is evidenced by the fact that on the high peaks of the Pamir mountains, fish skeletons are found, and the proof of this is the millions of remains of various mollusks that have not yet been studied on the ridges of our usual Karatau.
Публицистикалық	Қазақстанда коронавирустан тағы 4 адам қайтыс болды. Өлім нәтижелерін зерттеу жөніндегі комиссия бұл жағдайларды коронавирустық инфекциядан қайтыс болған деп тіркеуге шешім қабылдады.	В Казахстане умерло еще 4 человека от коронавируса. Комиссия по исследованию результатов смерти приняла решение зарегистрировать эти случаи как умершие от коронавирусной инфекции.	In Kazakhstan, 4 more people died from coronavirus. The death investigation Commission decided to register these cases as having died from a coronavirus infection.

Аудармашы бағдарламалардың өзіндік артықшылықтары мен кемшіліктері бар, 3-кестеде көрсетілген. Олардың әрқайсысының жұмысын талдау арқылы қазақ тілінен немесе қазақ тіліне аударғанда мүмкін болатын қателерді анықтай алдық.

3-кесте. **Аудармашылар жұмысының салыстырмалы сипаттамасы**

№	Аудармашылар	Кемшіліктері	Артықшылықтары
1	Google аудармашы	Үлкен сөйлемдерде кейбір сәйкессіздік бар, бірақ көп жағдайда ол өте аз.	Күрделі ғылыми мәтіндердің жақсы және сапалы аудармасы, кішігірім сәйкестік мәселелері
2	Яндекс аудармашы	Үлкен мәтіндерді аудару кезінде сөйлемдердің бөліктері үзіліп, аударылмайды	Өте үлкен сөйлемдерді, күрделі ғылыми мәтіндердің сапалы аударды
3	<u>Tridentsoftware</u>	Үлкен сөйлемдерде кейбір сәйкессіздік бар	Қысқа сөйлемдер мен сөздер үшін жақсы аударма нәтижелері.

4	Prof-translate	Аударма кезінде сөйлемдердің бір бөлігі қабылданбайды, ұзақ сөйлемдерді үйлестіруде қиындықтар байқалады	Қысқа сөйлемдер мен сөз тіркестер үшін жақсы аударма нәтижелері.
5	sozdik.kz	Үлкен ғылыми мәтіндердің аудармасы жақсы нәтиже бермейді	Сөздер мен қысқа сөйлемдер үшін жақсы аударма нәтижелері.
6	translate.zakon.kz	Үлкен сөйлемдерде кейбір сәйкессіздік бар	Қысқа сөйлемдер мен сөздер үшін жақсы аударма нәтижелері.
7	elim.kz	Аударма кезінде сөйлемдердің бір бөлігі қабылданбайды	Сөздер мен қысқа сөйлемдер үшін жақсы аударма нәтижелері.
8	Prompt	Үлкен ғылыми мәтіндерді аударуда өте жақсы нәтиже бермейді, көптеген сөйлемдер келісілмейді	Журналистік стильдегі қысқа сөйлемдердің жақсы аударма нәтижелерін көрсетеді

IV. Қорытынды

Тіл білу дәстүрлі түрде білімді адамдардың ерекшелігі болды. Бүгінгі таңда машиналық оқытудың озық технологиялары мен жоғары сапалы деректердің үйлесімінің арқасында автоматты аударма нәтижесі алынады, көп жағдайда адамның жасаған аудармасынан сапасы төмен емес. Машиналық аударма жақын арада қолмен аударманы алмастыра алмайды, өйткені бұл жасанды интеллект технологияларын қажет етеді, олардың ақыл-ой белсенділігі адамға тең болуы керек. Егер адамдар сөзді басқаша жаза бастаса, жүйе мұны жаңа мәтін енген бойда көреді. Аударма сапасын жақсарту мақсатында жүйе үнемі жаңартылып отырады және тексерулер жүргізіледі, бұл аудармашылардың жұмысын айтарлықтай жеңілдетеді және жеделдетеді. Машиналық аударманың артықшылығы – мәтіннің негізгі идеясын қысқа уақытта ұзақ қолмен аударуды талап ететін аудармасыз түсіну. Жасанды интеллект адам немесе одан жақсы аударма алатындай дамуға жеткенде, машиналық аударма қолмен қарағанда жақсы деп айтуға болады.

АЛҒЫС

Бұл жұмыс AR08052421 жобасы аясында, ҚР Білім және ғылым министрлігінің қолдауымен жүзеге асырылды.

ӘДЕБИЕТТЕР

- [1] Zhumanov Zh.M., Tukeyev U.A. Development of machine translation software logical model (translation from Kazakh into English language). Reports of the Third Congress of the World Mathematical Society of Turkic Countries, Volume 1 (June 30 – July 4, 2009) / Edited by Academician Bakhytzhan T. Zhumagulov. – Almaty: Қазақ университеті, 2009. – 356-363 p.
- [2] Tukeyev U., Zhumanov Zh., Rakhimova D. Features of development for natural language processing. In book “ICT – from theory to practice” edited by M.Milosz. Polish Information Processing Society, Lublin, 2010, 149-174 pp.
- [3] Tukeyev U., Rakhimova D. Augmented attribute grammar in meaning of natural languages sentences. The 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligent Systems, SCIS-ISIS2012, Kobe, Japan on November 20-24, 2012, 1080-1085 pp. (Индекс Скопус)
- [4] Тукеев У.А., Рахимова Д.Р., Байсылбаева К., Умирбеков Н., Оразов Б., Абақан М., Кызырканова С., Көпмағыналық бейнелеу кесте тәсілі негізінде орыс тілінен қазақ тіліне машиналық аудармасының морфологиялық анализбен синтезін құру. түркі тілдерін компьютерлік өңдеу. Бірінші халықаралық конференция: Еңбектері/ Астана: Л.Н.Гумилев атындағы ЕҰУ баспасы, 2013, 182-191.
- [5] Тукеев У.А. Разработка технологии машинного перевода на основе метода многозначных отображений для морфологически сложных языков. Труды 4-ой Международной научно-практической конференции «Информатизация общества», Астана, 2014, стр. 130-132.
- [6] Tukeyev, U., Milosz, M., Zhumanov, Zh. Finite-State Transducers with Multivalued Mappings for Processing of Rich Inflectional Languages. // Lecture Notes in Computer science. New trends in intelligent information and database systems (Vol. 598, pp. 271-280). Springer. 2015. (Индексировано в WoS, Scopus)
- [7] Tukeyev, U., Automaton models of the morphology analysis and the completeness of the endings of the kazakh language. Proceedings of the international conference “Turkic languages processing” TURKLANG-2015 September 17–19, Kazan, Tatarstan, Russia, 2015. pp. 91-100
- [8] Tukeyev U.A., Rakhimova D.R., Zhumanov Zh.M., Kartbayev A.Zh. Single state transducer model for Kazakh and Russian morphology // KazNU BULLETIN, Mathematics, Mechanics, Computer Science Series. – Алматы, «Қазақ университеті». – 2016. – №2 (89). – P. 110-117.
- [9] Tukeyev U., Sundetova A., Abduali B., Akhmediyeva Zh., Zhanbussunov N. Inferring of the morphological chunk transfer rules on the base of complete set of Kazakh endings // Lecture Notes of Artificial Intelligence (LNAI) vol.

9876, Computational Collective Intelligence, Part 2, Springer, 2016, pp.563-574 (Индексировано в WoS, Scopus)

[10] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic Regularities in Continuous Space Word Representations. The Association for Computational Linguistics. In HLTNAACL, p. 746–751(2013).

[11] Nal Kalchbrenner, Phil Blunsom. Recurrent Continuous Translation Models. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, p. 1700–1709(2013)

[12] Mikel L. Forcada and Ramon P. Neco Recursive Hetero-Associative Memories for Translation. International Work-Conference on Artificial and Natural Neural Networks, IWANN'97 Lanzarote, Canary Islands, Spain, p. 453-462(1997).

[13] Ilya Sutskever, Oriol Vinyals, Quoc V. Le Sequence to Sequence Learning with Neural Networks. arXiv:1409.3215v3 [cs.CL](2014).

[14] <https://ru.wikipedia.org/>

[15] How does Neural Machine Translation work? | SYSTRAN Blog: [Электронный ресурс]. 2016. URL: <http://blog.systransoft.com/how-does-neural-machine-translation-work/>

[16] eMPty Pages: A Deep Dive into SYSTRAN's Neural Machine Translation (NMT) Technology: [Электронный ресурс]. 2016. URL: <https://kv-emptypages.blogspot.ru/2016/09/a-deep-dive-into-systrans-neural.html>

[17] Колганов Д.С., Данилов Е.А. Обзор аналитической, статистической и нейронной технологий машинного перевода // Материалы X Международной студенческой научной конференции «Студенческий научный форум» URL: <http://scienceforum.ru/2018/article/2018009616>

"http://scienceforum.ru/2018/article/2018009616" (дата обращения: 20.05.2020).</p>

Рахимова Д., Турарбек А., Карюкин В., Карибаева А., Турганбаева А.

Обзор современных технологий машинного перевода для казахского языка

Резюме. В статье рассматриваются технологии машинного перевода, применение этих технологий для казахского языка, определение возможных ошибок машинного перевода, а недостатки и преимущества переводчиков.

Ключевые слова: машинный перевод, технологии машинного перевода, казахский язык, ошибки перевода, морфологический анализ.

УДК 004.8

N.S. Zhumakhan

(Almaty University of Power Engineering and Telecommunications Republic of Kazakhstan, Almaty
E-mail: nazym.sembekkyzy@gmail.com)

EYE BLINK DETECTION WITH OPENCV, PYTHON AND DLIB

Annotation. The article shows that the application of a real-time algorithm using the OpenCV and dlib libraries to determine eye blinking through facial landmarks is one of the most effective methods and is suitable for beginners in face recognition. When determining the level of eye openness, the results are sufficiently evaluated and the results are reflected in the article. Thus, the proposed algorithm evaluates the state of landmarks and obtains one scalar – the fundus ratio (EAR), which characterizes the detection of the source. Ultimately, the SVM classifier will identify eye blinking as an example of EAR values in a short time window.

Key words: eye blink detection, dlib, OpenCV, face recognition, algorithm, Python.

Н.С. Жұмахан

(Алматы энергетика және байланыс университеті, Қазақстан, Алматы
E-mail: nazym.sembekkyzy@gmail.com)

OPENCV, PYTHON ЖӘНЕ DLIB КӨМЕГІМЕН БЕТ БАҒДАРЛАРЫ АРҚЫЛЫ КӨЗДІҢ ЖЫПЫЛЫҚТАУЫН АНЫҚТАУ

Аңдатпа. Мақалада бет-әлпет бағдарлары арқылы көздің жыпылықтауын анықтауда OpenCV және dlib кітапханалары көмегімен нақты уақыт алгоритмін қолдану – тиімді тәсілдердің бірі және бет-әлпетті тану бойынша жаңадан үйренушілер үшін қолайлы екендігі көрсетіледі. Көздің ашылу деңгейін анықтауда жеткілікті түрде бағаланып, анықталған нәтижесі мақалада көрсетіледі. Осылайша, ұсынылған алгоритм бағдарлардың жағдайын бағалайды, бір скалярлық шаманы – көздің табылуын сипаттайтын көз түбінің коэффициентін (EAR) алады. Соңында, SVM классификаторы көздің жыпылықтауын қысқа уақыт терезесінде EAR мәндерінің үлгісі ретінде анықтайтын болады.

Түйін сөздер: көздің жыпылықтауын анықтау, dlib, OpenCV, бет-әлпетті тану, алгоритм, Python.