

«Сейфуллин оқулары – 16: Жаңа формациядағы жастар ғылыми – Қазақстанның болашағы» атты халықаралық ғылыми-теориялық конференциясының материалдары = Материалы Международной научно-теоретической конференции «Сейфуллинские чтения – 16: Молодежная наука новой формации – будущее Казахстана». - 2020. - Т.І, Ч.3 - С.207-209

ДЕРЕККӨЗДЕРДІ ЛИНГВИСТИКАЛЫҚ ӨНДЕУ АРҚЫЛЫ ҚАЗАҚ-АҒЫЛШЫН МАШИНАЛЫҚ АУДАРМАНЫ ЖЕТІЛДІРУ

Жумағалиева А.Ш.

Зерттеу барысында қазақ-ағылшын машиналық аудармасының артықшылықтары мен кемшіліктері қарастырылады. Зерттеудің басты мақсаты қазіргі заманғы машиналық аударма жүйелерінің көмегімен аударумен байланысты негізгі проблемаларды талдау болып табылады. Осыған байланысты біз компьютерлік аударманың сапасын қарастырдық, бағалау жүргізілді және машиналық аударманың типтік қателіктері мен бұрмаланулары анықталған. Жүргізілген талдауға сүйене отырып, компьютерлік аударманы оңтайландыру бойынша ұсыныстар берілді.

Сізге қажетті ақпаратты тез әрі сапалы алуға мүмкіндік беретін қазіргі заманғы ақпараттық технологиялар саласының жоғары дамуын ескере отырып, кейбір пайдаланушылардың интернетке жүгінуге және тиісті ақпаратты бір тілден екінші тілге аударуға деген ұмтылысы түсінікті. Интернет біздің өміріміздің ажырамас бөлігіне айналды. Қысқаша мәтінді немесе бір сөйлемді жедел аударуды қажет ететін, кәсіби аударма үшін аударма агенттігіне әрдайым жүгінуге болатын адам бола бермейді, ал басып шығарылған және тіпті электронды сөздіктерді қолдану шет тілінің грамматикасын білмейтін адам үшін қажет емес. Мүмкін, осы себептерге байланысты қазіргі уақытта онлайн-аудармашылар танымал, бұл мәтіндерді, файлдарды және веб-сайттарды бір тілден екінші тілге тез әрі жылдам аударуға мүмкіндік береді.

Зерттеу нысаны ағылшын тілінен қазақ тіліне аударатын машиналық аударма жүйесі. Зерттеу пәні лингвистикалық деректерді өндеуді қолдана отырып, қазақша-ағылшынша машиналық аударманы жетілдіру.

Зерттеу барысында эмпириалық зерттеу әдісі, зерттеудің теориялық деңгейіндегі қолданылатын әдістер, салыстырмалы талдау әдісі қолданылды.

Қазіргі уақытта онлайн-аудармашылар танымал, бұл мәтіндерді, файлдарды және веб-сайттарды бір тілден екінші тілге тез әрі жылдам аударуға мүмкіндік береді. Машиналық аударма (МА) – бұл жазбаша мәтіндерді бір табиғи тілден екіншісіне арнайы компьютерлік бағдарлама көмегімен аудару процесі. [1]. Интернеттегі мәтіндермен жұмыс істеуге және тез аударманы жүзеге асыруға мүмкіндік беретін машиналық аударма құралдары адамдар үшін байланыс құралы болып табылады. Қазіргі кезде барлық тілдерде бірдей дамыған автоматты аудармашылар жоқ. Машина аудармасының ғасырлық тарихы бар екеніне қарамастан (компьютерлік аударма сандық талдау машинасының жобасынан бастап жыл санау

жүргізеді Г. Бэббиджа XIX ғасырдың ортасынан бастап [1]), осы уақытқа дейін компьютерлік индустрияның бұл саласы жетілмеген күйінде қалып отыр.

19-20 ғасырдың ішінде халықаралық тілді құруға көп әрекет болды (эсперанто пайда болды) бірақ машиналық аударма жасау процесі тоқтатылды. 1933 жылы, Петр Смирнов-Троянский аударма үшін өз патентін ұсынғанда, зерттеулер үшін жаңа импульс пайда болды. Оның нұсқасы-бұл адам мен машинаның өзіндік ынтымақтастығы: бірінші редактор мәтіннің базалық сөздерін және оның синтаксистік функцияларын логикалық талдауды жасайды, машина бірінші редактордың қорытындысын аударуды жасайды, ал екінші редактор мәтінді мақсатты тілге бейімдейді.[2]

Зерттеудің негізгі мақсаты грамматикалық және лексикалық мәселелерді қарастыра отырып, қазақша-ағылшынша машиналық аударма сапасын арттыру және статистикалық үлгілер арқылы ағылшын-қазақ машиналық аударма сапасын жақсарту болып табылады. Аударманың статистикалық моделінің негізгі сипаттамалары қазақ тілін морфологиялық талдау әдісін қолдану есебінен жақсарады. Адамдар үнемі емле қателерін жасайды, және олардың өзіндік өрнегі бар, мысалы, смайликтер мен мәтіннің кейбір әшекейлерін пайдалана отырып кейбір тыныс белгілерін ерекше мақсатта қайталау немесе пайдалану. Содан кейін орфографиялық қателер машиналық аударма саласындағы ең үлкен мәселелердің бірі болып табылады. Бұл санинтернет пайдаланушылары жаңалықтарда, форумдарда, әлеуметтік желілерде жазған мәтіндермен айтарлықтай артуда. Қазақстанның веб-пайдаланушылары айтылуға ұқсас кейбір әріптерді басқаларына жиі ауыстыруы әбден мүмкін. Бұл орфографиялық қателер сөздерді теңестіруге айтарлықтай әсер етеді, сондай-ақ аударма мен тіл үлгілерінің сапасын төмендетеді. Кейбір бастапқы сөздер аударма үлгісі үшін белгілі емес болғандықтан машиналық аударма көмегімен аударылуы мүмкін емес. Мұндай белгісіз сөздердің мысалдары - дұрыс зат есімдер, әртүрлі морфологиялық формалар мен мәндер бар етістіктер. Морфологиялық сегменттеу сөзді оның құрамдас бөліктеріне бөлуге арналған. Қазақ тілі үшін бұл түбір есімдерінің, іс-әрекеттердің, етістіктердің және морфемдердің әртүрлі түрлерінің арасындағы сөздердің сегменттелуін қамтиды. Қазақ тілі үшін тағы бір ерекше проблема сөздердің қысқа түрлерін тану болып табылады. Кейбір жағдайларда беттердің екі түрі арасындағы негізгі айырмашылық олардың мағынасында болады. Қазақ тілі морфологиялық күрделі тіл болғандықтан, қолайлы қазақ морфологиялық сегменттеуін таңдау, машиналық аударма деректерін алдын ала өңдеудің өте маңызды кезеңі болып табылады. Бұл таңдау жазбаша аударма сапасына айтарлықтай әсер етеді. Сегменттеу схемасы оқытудың барлық жиынтықтары, күйге келтіру және тесттер арасында келісілуі тиіс. Көптеген морфемдер ағылшын жалғау ретінде жұмыс істейді: іс, зат есімдер және етістіктер әдетте қоса беріледі. Деректер жинағын алдын ала өңдеу бастапқы және мақсатты тілдердің арасындағы морфологиялық айырмашылықтарды төмендетуі

мүмкін екенін ескере отырып, бұл аударма үлгілері сөздердің тегістелуін жақсартуға және жүйенің өнімділігін арттыруға көмектеседі.

Қазақ тіліндегі машиналық аударманы зерттеуде кез келген басқа тілдердің арасындағы аударма кезіндегі проблемалар кездеседі, бірақ кейбір нақты сұрақтар бай агглютинативті табиғат сияқты тілмен тығыз байланысты. Машиналық аударманың негізгі мәселелерін бірнеше негізгі санаттарға бөлуге болады: сөздердің екі мағыналығы, терең синтаксистік және лексикалық айырмашылық, идиомалар мен фразеологизмде.

Журавски мен Мартин өзінің атақты кітабында аударманың күрделілігі тілдердің арасындағы іргелі айырмашылықтардың нәтижесі болып табылатындығын және ұқсас тілдердің арасында әртүрлі шығу тегі бар тілдерге қарағанда аудару оңай болуы мүмкін екенін түсіндірді. [3]

Қазақ тілін машинамен аударудың негізгі мәселелерін келесі санаттарға бөлуге болады: бір мағыналы емес, тілдік айырмашылықтар, адам қателіктері, диалектілердегі айырмашылықтар және техникалық проблемалар.

Талдау нәтижесінде деректер бізге толығымен автоматтандырылған аударма (адамнан тәуелсіз) бүгінгі күні мүмкін емес деп айтуға мүмкіндік береді. Қазақ тілі үшін тағы бір ерекше проблема сөздердің қысқа түрлерін тану болып табылады. Басқа түркі тілдері сияқты қазақ тілінің күрделі морфологиялық құрылымы, бай морфологиясы бар. Бірақ машиналық аударма нәтижелері мәтін сигналдық ақпарат ретінде қолданылған және мұқият редакциялауды талап етпейтін жағдайда мазмұнмен беттік танысу үшін пайдаланылуы мүмкін. Осылайша, біз оқытушыға ғылыми-техникалық мәтіндермен танысу, аудару үшін компьютерлік бағдарламаларды пайдалану кезінде олардың қауіптілігін сақтауды студенттерге ескерту қажет деп санаймыз. Бұл ретте олардың назарын машинамен жіберілген қателерді шебер және сауатты түзету және алдын ала болжауға тура келетініне аудару керек.

Ғылыми жетекші: Айтимова У.Ж. ф-м.ғ.к, аға оқытушы

Әдебиеттер тізімі

- 1 Марчук Ю. Н. Проблемы машинного перевода.- М.: Наука, 1983.- 201 с
- 2 <http://www.textologia.ru/yazikoznanie/interesnie-fakti-yazikoznania/jaziki-mira/kratkaya-istoriya-mashinnogo-perevoda/6230/?q=641&n=6230>
- 3 Jurafsky D. and Martin J. Speech and Language Processing: An Introduction to Natural Language Processing.-Prentice-Hall.-2000.-p.411.
- 4 <http://www.informatica.si/>