

УДК 81.25

МАШИНАЛЫҚ ӨНДЕУДЕГІ СЕМАНТИЗАЦИЯ МӘСЕЛЕЛЕРІ

Машенова М.Н., 2 курс, аударма ісі, А.Байтұрсынов атындағы Қостанай өңірлік университеті

Жабаева С.С., аға оқытушы, шетел тілі кафедрасының меңгерушісі, филология ғылымының кандидаты, Қостанай өңірлік университеті

Қазіргі заманғы көп тілді және мультимәдени әлемде тілдердің тоғысу мәселесі, қоғамдарды біріктіру бойынша тілдер саласындағы тиімді және өмірге бейімді бағдарламаларды іздеу ерекше өзекті болып табылады. Бүгінгі күні дүниежүзінде машиналық аударманы қолдану арта түсті. Қолданыс арта түскенімен оның дұрыс аудару сұранысы одан да бетер артты. Сонымен қатар бұның артуының себебі бірнеше проблемаларға және қиындықтарға әкеп соқты. Тізбектеп айтатын болсақ семантикалық, синтаксистік, мәдени-әлеуметтік тілдердің айырмашылықтары бұл проблемаларды айқынырақ көрсетеді және шешімді оданда қиындатады.

Машиналық өңдеу әлі күнге дейін толық зерттелмеген ғылым болып табылады. Машиналық өңдеудің негізі мақсаты бір тілден екінші тілге аудару. Оның мақсатына қарамастан аударма көркем әдебиеттегі ағылшын романы Гамлетті қазақ тілін аудару болсын, ғылыми мақаланы болсын. Қайсы бірін алсаңда анығы бір тілден екінші тілге тиянақты дұрыс аудару алу қабілеті керек. Яғни «Human Translations» адам аудармаларына сәйкес ойға қонымды аударма болуы тиіс. Машиналық өңдеудің аудару қабілеті өздігінен пайда болмайды ол түсінігін тек бір тілден екінші тілге ауыстыра алады.

Машиналық өңдеудің (аударманың) көп зерттелген және талқыланған екі түрі бар (Semantic Machine Translation) Семантикалық Машиналық Өңдеу және (Statistical Machine Translation) Статистикалық Машиналық Өңдеу.

Машиналық Өңдеудің екеуінің де механизмдерінің негізгі мақсаты трансформация арқылы түпнұсқадағы мәтінді екінші тілге конвертациялау. Конвертациялау нәтижесіндегі аударма мәтіні бір идеяны екі тілде теңбе-тең жеткізу болып табылады. Яғни бір тілдегі мәтінді келесі тілге аудару. Түпнұсқа мәтіні, басқа тілге аударылған мәтін. Мысалы, ағылшын тілі «I am writing», қазақ тілі «Мен жазып отырмын».

Статистикалық машиналық өңдеу режимідегіміз қарапайым параллель мәтіннің моделі, яғни сөйлем жұптары басқаларға қарағанда бір-бірінің аудармасы ретінде орын алуы ықтимал модель. Тиісінше, статистикалық машиналық аударма жүйесін құру үшін тілдік жұптарды жинау қажет аударма тіліне аудару үшін. (3)

Статистикалық машиналық өңдеу әдетте тілдің барлық ережелерін дамытады және декодтайды: жалғауларды, жіктік, шақтарды және т.б. Миллиондаған сөйлемдер талдау статистикалық машиналық аудармаға берілген. Кейін талдай отырып, негізгі өтпелі процестің негізін қалайтын мүмкіндіктердің үлкен базасы қалыптасады.

Статистикалық машина өңдеуінің қолданыста кең таралған түрі Google Translator. Бұл бағдарламаны 200 миллионға жуық адам күнделікті өмірде қолданады. Google Translator бағдарламасы қалай жұмыс істейді деген сұрақ туындайды. Google Translator бағдарламасының негізі іздеу жүйесінде. Яғни миллиондаған көптілді мәтіндердің сайттардан интернет арқылы іздеп табады да соған сәйкес келетін аударманы ұсынады. Айта кететін жайт, Google Translator бағдарламасын тұтынушыларына ерекше мүмкіндік береді. Ол берілген аударманы түзетуге, және ол ақпарат міндетті түрде сақталады. Бұл аудармаларды реттеу үшін мүлдем

**ҚҰЗЫРЕТТІЛІК БІЛІМ: МОДЕЛЬДЕР,
ӘДІСТЕР, ТЕХНОЛОГИЯЛАР
КОМПЕТЕНТНОСТНОЕ ОБРАЗОВАНИЕ:
МОДЕЛИ, МЕТОДЫ, ТЕХНОЛОГИИ**

тіл маманы болудың қажеті жоқ. Әр тұтынушының алатын орны маңызды роль атқарады. Тіпті бағдарламаның жүйесіне арнай баптаулар немесе жаңартулар керек еместі Машиналық өңдеу келесі тілдерге, олар: ағылшын тілі, орыс тілі, француз және неміс тілдеріне нәтижелі аударма жасайды. Алайда, Машиналық өңдеу пайда болғаннан бері семантикалық модель қажеттілі «Human Translations» адам тәржімасына қол жеткізу деген тұжырым айтылып келеді /Weaver 1955/. Семантиканың «Бекіту» екенін дәлелдеу Машиналық өңдеуде өте кең таралған; ол сізді аударманы тиянақты етіп түсіндіріп аудару кең таралған сенімін білдіреді және кейбір тәсілдерді басқалардан қорғау үшін жиі қолданылады /Сантос, 1992/.

Осыған қарамастан, семантикалық машиналық өңдеу іс жүзінде қиын, кейде тіпті шығармашылық процесс болып табылады бұған Statistical Machine Translation-ге қарағанда көп тіл мамандары мен бағдарламашылар қатысады. Семантикалық машиналық өңдеу әліжетілмеген құбылыс. Қазіргі уақытта семантикалық машиналық өңдеу алдында тұрған негізгі мәселелері келесідей:

Көп мағыналы сөз

He deposited money in the bank account with a high interest.

Sitting on the bank of Thames, a passing ship piqued his interest.

Көптеген тілдерде bank және interest сөздері үшін әр түрлі аудармалар болады. Аудару мақсатында бұл сөйлемдерде адам ақыл-ойы елегінен өткізіп мәтінді ескереді. Бірақ компьютер мұны істей алмайды.

Сөз байлығы

Машиналық өңдеу кезінде керекті сөзді тауып ойды тиянақты түпнұсқаға сәйкестендіре алмауында.

Идиомалық тіркестер

Машиналық өңдеудің сөз байлығы болмағандықтан, идиомалық тіркестерді түсіне алмайды

Мысалы: Raining cats and dogs

Бұл идиома қазақ тіліне сөзбе-сөз аударудың ешқандай мағынасы болмайды.

Стильді сақтау

Машиналық өңдеу стильді сақтамайды. Мысалы: When I got up, the morning had already been broken. Бұл сөйлемді аудару, бастапқы мәтін стильді жоғалту мүмкіндігі өте үлкен.

Корреферентілік

Мысалы: When I visit my aunt, I can't help myself playing with my little cousin.

Нағашы еркек пе, әлде әйел ме?

6. Грамматикалық сөйдемдер

Машинаны өңдеу кезінде аударма процесін күрделендіретін тағы бір маңызды фактор аударма әлеуметтік және мәдени мәселелерге қатысты, мысалы, қазақтың сөзі қалың мал тән тек қазақтар үшін және оның ағылшын тіліне сөзбе-сөз аудармасы мағынасы болмайды.

Тимочко мағыналық тапсырмалар әлеуметтік және мәдени факторларға тәуелді деп тұжырымдайды, өйткені мағыналық тапсырмаларды ажырату үшін маңызды дәлелдер мағыналық емес: екі бірдей ақылға қонымды мағыналық тағайындаулар, егер ол өтелген болса, әр түрлі шындық мәндерін тағайындай алады прагматиканың әртүрлі теориялары. Оның пікірінше, маңыздысы тек мәтін ғана емес, сонымен қатар спикерлердің жеке көзқарастары, олардың сенімдері. Сондықтан табиғи тілдер арасында аударма жасау керек әр тілде сөйлеушілердің қоғамы мен сенімдері туралы белгілі бір білімге ие болу қажет /Тимочко 1978/ Дывик екі тілдің семантикасын бейнелейтін жағдай схемаларын ұсынады, олардың екі тіл үшін бірдей болуы шарт емес екенін мойындай отырып семантикалық машина өңдеу армян, фин, серб және т.б. сияқты шағын тілдерге өте пайдалы.

**ҚҰЗЫРЕТТІЛІК БІЛІМ: МОДЕЛЬДЕР,
ӘДІСТЕР, ТЕХНОЛОГИЯЛАР
КОМПЕТЕНТНОСТНОЕ ОБРАЗОВАНИЕ:
МОДЕЛИ, МЕТОДЫ, ТЕХНОЛОГИИ**

Машиналық аударманы пайдаланушылар машиналық аударма әлі жетілмеген екенін түсініп, шешуші рөл атқарады. Ақылды пайдаланушы, сана елегінен өткізіп мақсатты мәтіннің жалпы мағынасын түсінеді. Ана тілін жетік білген, аударма процесінде жіберілген әрбір қате айқын болып табылады. Мысалы: Бұл үлкен үй.

Осы сөйлемді ағылшын тіліне аудару кезінде көптеген нұсқаларға кезігуіміз мүмкін. Адамдар ақылды және адамның ақыл-ойы сөздерді модельдеуге және талдауға қабілетті, біз ең жақсысын таңдай аламыз.

1. The house is big good
2. House big is the bad, but still understandable
3. The house is XXL worst, not understandable and acceptable at all.

Бұл кезеңде грамматикалық тұрғыдан қарағанда біртектес нұсқа да күнделікті аударманы орындау мысалға алсақ Қытай немесе Португалия ақпарат агенттіктерінің тақырыптарын түсіну үшін қанағаттанарлық болады. Машиналық өндеудің талаптар әр түрлі: бір адам аударманы көзіне мүмкіндігінше сенімді етіп аударуды қалайды, екіншісі дұрыс естілетін аударманы қалайды. Эдинбург университетінің Профессоры Филлип Коун бір рет өзінің дәрістері кезінде «мақсатты мәтін әрқашан жақсы естілмеуі немесе жақсы тілде жазылуы мүмкін, бірақ ақылды пайдаланушы оны түсінеді».

Қалай семантикалық машиналық өңдеу жұмыс істейді? Графикалық құрылымдық мағынаны білдіру семантикалық машина өндеудің негізі болып табыла. Қолданыстағы мәтін өндеудің семантикалық машиналық аударма мен статистикалық аударманың негізгі процестерінің айырмашылығы айрықша. Семантикалық машиналық өндеуінің жағдайында аударма процесі екі деңгейден тұрады. Бірінші деңгей – төл мәтінді аралық деңгейге айналдыру: график құрылымды ұсыну мағынасы. Екінші деңгей – аралық кезеңнің мақсатты мәтінге айналуы. Бұл аралық кезеңді бағдарламашылар тіларалық деп атайды. Төл мәтін, тіларалық кезең, мақсатты мәтін. Мысалы, ағылшын тілі, тіларалық кезең, қазақ тілі.

Әрбір машиналық өңдеу процесінің нақты жоспары бар. Шындығында, бұл процесс мүмкін болатындай қарапайым емес. Келесі төмендегі графикте көрсетілгендей көптеген басқа әрекеттерді қажет етеді.



Тіларалық дегеніміз не? Машиналық өндеудің алғашқы сәтті әрекеттері қолмен енгізудің үлкен кестелеріне сүйенгін ережелер. Бұл жүйелер адамға жоғары мамандандырылған білім қажет етеді. Сирек ерекшеліктерді қоспағанда, бұл жүйелер не аударылған, не тіл аралық. Аудармаға негізделген жүйелерде сөйлемдер түпнұсқа тілде талданды, ереже-

**ҚҰЗЫРЕТТІЛІК БІЛІМ: МОДЕЛЬДЕР,
ӘДІСТЕР, ТЕХНОЛОГИЯЛАР
КОМПЕТЕНТНОСТНОЕ ОБРАЗОВАНИЕ:
МОДЕЛИ, МЕТОДЫ, ТЕХНОЛОГИИ**

лер осы талдаудың түрленуін жетілдіру мақсатында сөйлемді талдап, содан кейін мақсатты мәтінді құруға болады /Андреас 2012/.

Неліктен семантикалық машиналық өңдеуді қолдану керек? Машиналық аудармамен байланысты барлық қиындықтарға қарамастан, оны миллиондаған адамдар кеңінен қолданады. Біріншіден, шет тілінде жазылған мәтін туралы жалпы түсінік алуға көмектеседі. Аудару кезінде жапон тілінен ағылшын тіліне төл мәтінді, мақсатты мәтін аралас ағылшын тілінде болады, бірақ оның тиянақты ойы түсінікті болады. Екіншіден, интернет арқылы қарым-қатынас немесе сұхбат кезінде бұл қатысушылар сөйлеген кезде көмектеседі. Мысалға чаттар, жедел хабаршылар, бейресми хаттар және т.б. Семантикалық машиналық өңдеу – бұл көп ақша жұмсауды және көп лингвисттерді қажет ететін күрделі жүйе, статистикалық машиналық аудармаға қарағанда. Тағы бір қиындық тіл маманының субъективті көзқарасымен байланысты.

Бұл процесс ұзақ уақытты алады: үшін жоғары сапалы семантикалық машиналық өңдеу, мәліметтер базасын құру 10 жылдан астам уақытты алады. Бірақ бұған қарамастан ғалымдардың барлық күш-жігерінің ішіннен профессор Роуз Смит қазіргі кезде ешкім зерттеуші сияқты ойламайды деп санайды. 1950 жылдары күрделі мәтіндерді осындай жоғары сапалы машиналық аударма қысқа және орта мерзімді перспективада қол жетімді мақсатқа айналды.

Пайдаланған әдебиеттер тізімі

1. Dyvik, The PONS Project Features of a Translation System, Bergen, 1990
2. J.Andreas «Towards Semantic Machine Translation, University of Columbia», 2012
3. M.R. Costa-Jussa «An Overview of the Phrase-based Statistical Machine Translation Techniques, Cambridge University Press», 2012
4. R.Smith «English Today», 2001
5. D.Santos «Semantics and (Machine) Translation», 1992
6. S.Bird, D.Chiang «Machine Translation for Language Preservation», 2012
7. T.Тумoczko «Translation and Meaning», 1978
8. Weaver, Translation. In Machine translation of Languages, volume 14, MIT Press, Cambridge, MA., 1955
9. www.darpa.mil

УДК 321.7:070

**ТЕМА КАК ЖАНРООБРАЗУЮЩИЙ ФАКТОР
(НА ПРИМЕРЕ ТЕЛЕКАНАЛА КОФ АО РТРК «ҚАЗАҚСТАН» «QOSTANAY»)**

Абжанова Р.Р., 4 курс, 5В050400 – журналистика, институт экономика и права, Костанайский региональный университет им. А.Байтурсынова

Машикова С.Н., ассоциированный профессор кафедры журналистики и коммуникационного менеджмента, доктор PhD, Костанайский региональный университет им. А.Байтурсынова

В статье дается определение понятию «Журналистская тема», рассматривается связь между жанром и тематической составляющей телевизионного контента, с целью выяснения с какими проблемами сталкиваются журналисты при создании телевизионного