

ЭЖК 811.512.122'322.3

А. БАЙМЕНШИН<sup>1</sup>, С. ҚҰЛМАНОВ<sup>2</sup>,

Мемлекеттік тілді дамыту институты<sup>1</sup>,  
 А. Байтұрсынұлы атындағы Тіл білімі институтының аға ғылыми қызметкері,  
 филол.ғ.к., доцент<sup>2</sup>

## МАШИНАЛЫҚ АУДАРМА ЖҮЙЕЛЕРІНІҢ АУДАРМА ЖАСАУ ЖӘНЕ МОРФОЛОГИЯЛЫҚ ТАЛДАУ ПРИНЦИПТЕРІ ЖАЙЫНДА

### Аннотация

Мақалада өндірістік машиналық аударма жүйелерінің мақсаты мен оларға қойылатын талаптар, өндірістік және тәжірибелік машиналық аударма жүйелерінің айырмашылығы, ісқағаздарын қазақ тілінен орыс тіліне, орыс тілінен қазақ тіліне автоматты аударатын Moses программасының жадына грамматикалық сөз таптары белгілерін қою, қазақ тілінен орыс тіліне және керісінше орыс тілінен қазақ тіліне автоматты аударма жасаудағы іс-қимылдар алгоритмін Moses программасының жұмыс істеу принципіне үйлестіру механизмдері көрсетіледі.

**Тірек сөздер:** аударма, мәтін, ісқағаздары, машиналық аударма, морфологиялық талдау, сөзформа, пайдаланушы, сөз таптары, іс-қимыл алгоритмі.

**Ключевые слова:** перевод, текст, деловые бумаги, машинный перевод, морфологический анализ, словоформа, пользователь, части речи, алгоритм действий.

**Keywords:** translation, text, official papers, machine translation, morphological analysis, word form, the user of the speech, the algorithm actions.

Қолданбалы лингвистикада, жалпы машиналық аудармада морфология әдетте мәтінді автоматты өңдеудің нақты жүйелерін құруда шешуге тиіс практикалық міндеттерге қатысты оңайлатылып модельденеді. Машиналық аудармада морфологиялық талдау жасау сөзформаларды ақпарат алу мақсатында өңдеу дегенді білдіреді.

Өндірістік машиналық аударма жүйелерінің мақсаты мен оларға қойылатын талаптарды төмендегіше көрсетуге болады:

1. Өндірістік машиналық аударма жүйелері шет тілін меңгермеген мамандардың ақпараттық қажеттілігін қанағаттандыру үшін мәтіндерді бір тілден екінші тілге автоматты немесе жартылай автоматты аударуға арналады.

2. Өндірістік машиналық аударма жүйелері айтарлықтай мол көлемдегі мәтіндерді қамтиды және олар көптеген пайдаланушылардың қолдануына арналады.

3. Өндірістік машиналық аударма жүйелері дәстүрлі аудармамен салыстырғанда уақытты ұтуды қамтамасыз ету үшін аударманы аса жоғары жылдамдықта жасауды қамтамасыз етуге тиіс.

4. Өндірістік машиналық аударма жүйелері арқылы алынатын аударманың құны дәстүрлі аударманың құнымен салыстырғанда төмен болуға тиіс.

5. Өндірістік машиналық аударма жүйелері арқылы жасалатын аударманың сапасы жоғары болуға тиіс.

6. Өндірістік машиналық аударма жүйелерінде мәтінде кездесетін барлық грамматикалық құбылыстар қамтылуға тиіс.

7. Өндірістік машиналық аударма жүйелері арқылы аударылған мәтіндерге аздаған редакциялау жүргізілетіндей болуға тиіс.

8. Өндірістік машиналық аударма жүйелері пайдалануда барынша жеңіл болуға тиіс [1, 18].

Өндірістік машиналық аударма жүйелерінен айырмашылығы модельдік (тәжірибелік) машиналық аударма жүйелері тәжірибелік сипатта болады. Сондықтан:

1. Тәжірибелік аударма жүйелері аударма қызметінің қандай да бір аспектілерін модельдеуге немесе қандай да бір лингвистикалық құбылыстарды тексеруге арналады.

2. Тәжірибелік аударма жүйелері программалау тілдерінің базаларында әзірленеді.

3. Тәжірибелік аударма жүйелерін әзірлеу құны қандай да бір тілдің ерекшелігіне байланысты әртүрлі болуы мүмкін.

4. Тәжірибелік аударма жүйелерін қолдану арқылы жүргізілетін лингвистикалық тәжірибелер шағын көлемдегі мәтіндерді қамтиды.

5. Тәжірибелік аударма жүйелерінің сөздігі мен грамматикасы зерттеу жүргізуге қажетті әрі жеткілікті шағын көлемде ғана жүргізіледі.

6. Тәжірибелік аударма жүйелері әдетте бір ғана данада жасалып, тираждалмайды және басқа жерге орнатуға берілмейді.

7. Өндірістік аударма жүйелерін әзірлеуде тексерілген және мақұлданған әдістер мен тәсілдер қолданылса, тәжірибелік аударма жүйелері негізінен қандай да бір лингвистикалық құбылысты тексеру немесе техникалық міндетті шешу үшін әзірленеді.

8. Тәжірибелік аударма жүйелері арқылы жасалатын аударманың сапасы жүйеге енгізілген лингвистикалық модельдің барабарлық деңгейіне байланысты болады.

9. Өндірістік аударма жүйелерін жүйенің өзіндік программалық компонентінен айтарлықтай асып түсетін кең программалық сервиске ие болса, тәжірибелік аударма жүйелері әдетте тәжірибе жасау мен оның нәтижелерін бағалауға қажетті әрі жеткілікті сервистік қызметке ие болады.

10. Өндірістік аударма жүйелерінен айырмашылығы тәжірибелік аударма жүйелері технологиялық қанегізделмейді, мұнда тәжірибенің және/немесе оның мазмұнының сипатын өзгерту оңай болады [1, 19].

Программаның сөздік қоры шағын болғанда оған барлық талап етілетін сөзформаларды енгізіп, морфологиялық талдаусыз аударма жасауға да болады. Алайда зерттеудің негізгі нысаны морфология болса, зерттеуші барлық тілдік құбылыстарды: сөз түрлендіруші, форма тудырушы, сөзжасамдық және т.б. тиісті модельдерді енгізуге міндетті.

Өндірістік машиналық аударма жүйелеріне арналған морфологиялық модельдер практикалық тұрғыда пайдалану шарттарымен байланысты төмендегідей бірқатар талаптарға жауап беру керек:

1. Өндірістік машиналық аударма жүйелеріне арналған морфологиялық модельдер мәтінді (ақпаратты) өңдеудің жоғары жылдамдығын қамтамасыз етуге тиіс (яғни морфологиялық талдау мен синтездеу кезеңдері аудармаға жұмсалатын уақыттың аз ғана бөлігін алуға тиіс).

2. Өндірістік пайдалану жағдайларында морфологиялық талдау шағын жүйесі алдын ала таныс емес мәтінмен «жұмыс істей алуға» және қателіктерді «жөндей білуге» тиіс.

3. Морфологиялық компоненттің құрылымы мен ұйымдасуы жүйенің сөздік қорын толықтыру әзірлеуші-лингвистің аз күш жұмсауын, ал пайдаланушының арнаулы білімінің аса жоғары болмауын талап ету үшін барынша қарапайым болуға тиіс. Бұл лингвистикалық, программалық және техникалық құралдар жиынтығымен қамтамасыз етілетін өндірістік машиналық аударма жүйелерін үшін аса маңызды.

4. Келесі талап – талдау мен синтездеудің толықтығы. Морфологиялық талдаудың толықтығы морфологиялық компоненттің нақты мәтіндерде кездесетін барлық формаларды талдау қабілетін білдіреді. Морфологиялық синтездеудің толықтығы берілген морфологиялық ақпарат бойынша кез келген талап етілетін сөз формасын жасай алу қабілетін білдіреді.

5. Морфологиялық талдау мен синтездеудің жылдамдығын арттыруға бөлінетін сөз компоненттерінің санын азайту септігін тигізеді. Сөзформа бөлінетін бөлік екеуден аспағанда аударма тез жасалады [1, 19].

Лингвистикалық ақпаратты екі түрде: лингвист басып шығарған бастапқы түрде және машиналық жүйенің жадында сақтаған жөн. Ақпаратты толықтыру және түзету бойынша барлық операциялар стандартты редакциялау рәсімдерінің көмегімен бастапқы массивтерде орындалады. Түзетілген және тексерілген ақпарат машиналық аударма жүйесіне оның сервистік программалары арқылы енгізіледі.

Жоғарыда аталған талаптарға талдай жасай келе, төмендегідей тұжырымдар шығарылды:

1. Өндірістік және тәжірибелік машиналық аударма жүйелеріне қойылатын талаптарда айырмашылықтар болады.

2. Машиналық аударма жүйелерін пайдалану және толықтыру технологиясы пайдаланушы өздігінше қолдану үшін барынша қарапайым болуға тиіс.

3. Сөздік мақалаларды құру кезінде сөз түрлендіру кластарын анықтау айтарлықтай қиындық тудыратындықтан, кодталатын сөздердің грамматикалық қасиеттері нақты анықталуға тиіс.

4. Сөз түрлендіру топтары есте сақтауға жеңіл, шағын, көрнекі белгілермен белгіленуге тиіс.

5. Сөз түрлендіру топтарының жүйесі мен олардың мнемоникалық белгілері сөздікпен байланысты әзірленуге тиіс.

6. Өндірістік машиналық аударма жүйелеріне арналған морфологиялық модельдерді әзірлеу оның программалық іске асырылуын есепке алынып жүргізілуге тиіс.

Жоғарыда аталған талаптарды ескере отырып, біз ісқағаздарын қазақ тілінен орыс тіліне, орыс тілінен қазақ тіліне автоматты аударатын Moses программасының жадына орыс тіліндегі грамматикалық сөз таптарын төмендегіше белгіледік:

С - существительное

П- прилагательное

Ч - числительное

М - местоимение

ГП - глагол переходный

ГН - глагол непереходный

СЗ - союз

Н - наречие

М - междометие

П - подражательное слово

МС - модальное слово.

Ал қазақ тіліндегі грамматикалық сөз таптары төмендегіше белгіленді:

З-зат есім

Сын-сын есім

Сан - сан есім

Ес - есімдік

Ет - етістік

Ш - шылау

Ү - үстеу

О - одағай

Ел - еліктеуіш

МС - модаль сөз.

Аударма процесінде орыс тіліндегі сөйлемді қазақ тіліне аударғанда төмендегідей іс-қимыл алгоритмі қолданылады:

1) бастауыш пен баяндауыш тобын табу керек;

2) бастауыш тобынан анықтауышты табу керек;

3) анықтауышты аудару керек;

4) анықталатын сөзді аудару керек;

5) анықтауышты анықтайтын сөздің алына қою керек;

6) баяндауыш тобынан толықтауыш пен пысықтауышты табу керек;

7) осы тұрлаусыз мүшелерді аудару кезінде олардың тұрақты сөз тіркестерімен, аналитикалық құрылымдарымен берілуіне назар аудару керек;

8) толықтауыш пен пысықтауышты олардың коммуникативтік маңыздылығына қарай орналастыру керек;

9) баяндауышты аударып, сөйлемнің соңына орналастыру керек;

10) бастауыш тобын да, баяндауыш тобын да аудару кезінде сөйлемнің соңынан басына қарай аударылады.

Ал қазақша сөйлемді орыс тіліне аудару кезіндегі іс-қимыл алгоритмі біршама басқаша:

1) бастауыш пен баяндауыш тобын табу керек;

2) осы топтардың коммуникативтілік маңыздылығының деңгейін анықтап, соған сәйкес олардың сөйлемдегі орнын анықтау керек;

3) бастауыш тобынан анықтауышты және анықталатын сөзді табу керек;

4) анықтауыштың «жайылу» дәрежесін анықтау керек, аударманың оңтайлы нұсқасын табу керек, содан кейін ғана қазақ тіліндегі анықтауыштың сөз табын білдірілу сипатына байланысты оның анықталатын сөзге қатысты орнын анықтау қажет;

5) баяндауыш тобында сөйлемнің тұрлаусыз мүшелерін олардың коммуникативтілік маңыздылық дәрежесіне байланысты орналастыру керек. Егер баяндауыш тобында анықтауыш болса, 4-тармаққа сәйкес іс-қимыл жасаймыз[2, 344].

Аударма процесінде қазақ тілінен орыс тіліне және керісінше орыс тілінен қазақ тіліне автоматты аударма жасауда жоғарыда аталған тәжірибелік машиналық аударма жүйелеріне қойылатын талаптар ескеріле отырып, аталған іс-қимылдар алгоритмін Moses программасының жұмыс істеу принципіне үйлестірдік.

#### ӘДЕБИЕТ

1 *Минтусова О.В.* Моделирование морфологического компонента промышленных систем машинного перевода: автореферат канд.дис. –10.02.19. –Москва, 1990. –32 с.

2 *Ахметжанова З.К.* Сопоставительное языкознание: казахский и русский языки. –Алматы, 2005. –408 с.

Байменшин А.

Институт по развитию государственного языка

*Кулманов С.К.*

Старший научный сотрудник Института языкознания им. А. Байтұрсынова, канд.филол.наук, доцент

#### **О ПРИНЦИПАХ ПЕРЕВОДА И МОРФОЛОГИЧЕСКОГО АНАЛИЗА, ИСПОЛЬЗУЕМЫХ В МАШИНЫХ СИСТЕМАХ ПЕРЕВОДА**

##### **Резюме**

В статье рассматриваются цели промышленных машинных систем перевода и требования, предъявляемые к ним, особенности промышленных и экспериментальных машинных систем перевода, вопросы внедрения условных обозначений грамматических частей речи в память программы Moses, вопросы соотношения алгоритмов действий при переводе с механизмами работы программы Moses.

*Baymenshin A.*

Institute for Development of State Language

*Kulmanov S.K.*

Senior researcher at the Institute of Linguistics A. Baitursynov,

kand. filol. nauk (PhD), Associate Professor

#### **ON THE PRINCIPLES OF TRANSLATION AND MORPHOLOGY ANALYSIS USED IN THE MACHINE TRANSLATION SYSTEMS**

##### **Summary**

The article addresses the objectives of industrial machine translation systems and requirements to them, especially industrial and experimental machine translation systems, introduction legend grammatical parts of speech in the program memory Moses, questions correlating algorithm action in translation mechanisms of the program Moses.