

**У.А. ТУКЕЕВ, Д.Р. РАХИМОВА, К. БАЙСЫЛБАЕВА, Н. УМИРБЕКОВ,
Б.ОРАЗОВ, М. АБАҚАН, С. КЫЗЫРКАНОВА.**

Әл Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан

КӨПМАҒЫНАЛЫҚ БЕЙНЕЛЕУ КЕСТЕ ТӘСІЛІ НЕГІЗІНДЕ ОРЫС ТІЛІНЕН ҚАЗАҚ ТІЛІНЕ МАШИНАЛЫҚ АУДАРМАСЫНЫҢ МОРФОЛОГИЯЛЫҚ АНАЛИЗБЕН СИНТЕЗІН ҚҰРУ

КІРІСПЕ

Машиналық аудару жүйесі күрделі және өте ауқымды болып табылады: құрамына екі тілді және одан да көп сөздіктерді қосады, ол қажетті грамматикалық ақпаратпен жабдықталған (морфологиялық, синтаксистік және семантикалық), ол эквивалентті, нұскалық және трансформациялық аударма сәйкестіктерін, сонымен қатар грамматикалық талдаудың алгоритмдік құралдарын жіберуді қамтамасыз ету үшін қажет. Осы мақалада орыс тілінен қазақ тіліне және керсінше машиналық аудармадағы морфологиялық талдау мен синтездің жеке жағдайы қарастырылады.

Сөздер өзара байланыспай, бір-бірімен тіркеспей тұрғанда ойды және нақты мағынаны білдіре алмайды. Сөздердің бір-бірімен байланысуы және мағынаның толық болуы аффикстердің жалғауы арқылы іске асады.

Қазақ тіліне де, орыс тіліне де жалғанатын жұрнақтар мен жалғаулардың өзіндік жалғану реті мен жүйесі бар [1]. Бұл тілдің ішкі заңдылығына бағынатын күрделі тарихи процесс. Түбір мен қосымшалардың мағыналарының берілуінің де өзіндік тәртібі болады. Қазақ тіліндегі жалғаулар 4 топқа бөлінетін болса: көптік, септік, жіктік, тәуелдік, орыс тіліндегі жалғаулар «род, число, падеж (септік)» сияқты грамматикалық мағынада бөліне алады. Соның ішінде орыс тілі мен қазақ тіліндегі септік жалғауларға қатысты мәселерді қарастырсақ. Орыс тіліндегі «книга» сөзін септеп көрейік: именетильный падеж- книга, родительный падеж - книги, дательный падеж - книге, венительный падеж - книга, творительный падеж - книгой, предложный падеж - о книге. Бұл жерде орыс тілінен қазақ тіліне машиналық аударма кезінде қиындық туындайды себебі, екі септікте де «е» жалғауы кездеседі және «и» жалғауы «родительный» септігінде және атау септіктің көпше түрінде де кездесе алады. Мысалы: «книги нет на месте» және «я забрала все свои книги». Бірінші жағдайда бір ғана кітап жайлы және сол кітаптың орнында жоқтығы жайында сөз қозғалса, екінші жағдайда бірнеше кітап және соларды алып кеткендігі жайлы баяндалады. Бұның мағынасын біз жеңіл түсінгенімізбен, машинаның түсінуі екіталай.

Машиналық аудармадағы талдау тізбектелген және өзара байланысатын процесс болып саналады. Морфологиялық талдаудың нәтижесі синтаксистік талдауға әсерін тигізеді, ал морфология пен синтаксистік мәтіннің семантикасына ықпал тигізеді. Сондықтан мәтін талдауының бастапқы кезеңіне дұрыс көңіл бөлуді қажет етеді.

Біздің зерттеу жұмысымыздың негізгі бағыты: қазақ жалғауларының (аффикстерінің) түрлері мен байланыстардың атқаратын ролі және мәтіннің аудармасына тигізетін әсерін көрсету. Орыс-қазақ машиналық аудармадағы морфологиялық талдаудағы қазақ тілінің жалғауларын (аффикстерін) қолайлы жүйеге түрлендірудің әдісін ұсыну болып табылады.

1 КӨПМАҒЫНАЛЫҚ БЕЙНЕЛЕУ КЕСТЕ ТӘСІЛІ

Машиналық аудармадағы негізгі мәселердің бірі көпмағыналық болып саналады. Оған қатысты көптеген әдістер бар. Қазіргі қолданыста ең көп тарағаны статистикалық және ережелер негізіндегі әдістер болып табылады. Бірақ олардың өзіндік артықшылықтары мен кемшіліктері бар.

Бұл жұмыста біз табиғи тілдерді бір біріне машиналық аударма күрделі мәселелерін шешу жолдарын іздеу бағытында келесі пікір(ұйғарым) енгізіп жатырмыз. Әр сөзді тек екі бөлімнен құрылады деп санаймыз – негізден және жалғаулардан. Осы ұйғарымды компьютерлік көрсетуілінде біз **көпмағыналық бейнелеу кесте** түрінде қолдануын ұсынып жатырмыз. Көпмағыналы бейнелеу теория соңғы 10-15 жылдары қатты дамыған [2,3]. Машиналық аударма бағытында **бейнелеу** тәсілі қолданған мысалдары бар [4,5], бірақ **көпмағыналы бейнелеу** тәсілдерін қолданғанын қарастыра алмадық.

Көпмағыналық бейнелеу кестелер машиналық аударма процесін жүйректендіруге және жүйелендіруге ықпал тигізеді.

Машиналақ аударманы жасау барысында деректер қорына көпмағыналық бейнелеу кесте құру тәсілімен тілдер сөздігі мен қосалқы элементтер (жұрнақ, жалғау) кестелері құрылады. Қолайлық үшін орыс-қазақ электрондық сөздігін сөз табтарының түрлеріне сай кестелерге бөлуді ұйғардық. Электрондық сөйлем сөздігі тек қана сөйлем негізінен (түірінен) құралады. Және кестелерге әр сөз табына сәйкес қасиеттерді қосалқы атрибуттар ретінде енгіземіз. Кестелердегі атрибуттар қарапайым сандар ретінде белгіленеді. Мысалы: орыс тілдегі зат есімге "род", "число", "склонение"; қазақ тілінің етістіктеріне уақыт шағы, жақ және т.с.с.

Сөздер өзара байланыспай, бір-бірімен тіркеспей тұрғанда ойды және нақты мағынаны білдіре алмайды. Сөздердің бір-бірімен байланысуы, тіркесуі аффикстердің жалғауы арқылы іске асады. Грамматикада сөздер байланысы негізгі бес түрге (қиысу, меңгеру, матасу, қабысу, жанасу)бөлінеді, соның үшеуі (қиысу, меңгеру, матасу) жалғаулар негізінде құрастырылған. Соған қарағанда сөйлемдегі сөздердің мағыналық байланысы аффикстерге тәуелді.

Қазақ тіліне де, орыс тіліне де жалғанатын жұрнақтар мен жалғаулардың өзіндік жалғану реті мен жүйесі бар. Бұл тілдің ішкі заңдылығына бағынатын күрделі тарихи процесс. Түбір мен қосымшалардың мағыналарының берілуінің де өзіндік тәртібі болады.

Қазақ тіліндегі жалғаулар (аффикстер) 4 топқа бөлінетін болса: көптік, септік, жіктік, тәуелдік. Орыс тіліндегі жалғаулар «род, число, падеж (септік)» сияқты грамматикалық мағынада бөліне алады. Жалғаулардың сөз құрамына және құрылуына әсер ететіні мәлім. Орыс тілінен қазақ тіліне машиналық аударманы жасау барысында морфологиялық талдауда және синтезде көптеген мәселелерге және қиыншылықтарға тап болдық. Машиналық аудармаға морфологиялық талдауы мен синтезі үшін аффикстер көпмағыналық бейнелеу кестелер тәсілімен құрылады. Оларға да сәйкесінше қасиеттік атрибуттар енгізіледі.

Мысалы, орыс тілі етістік сөздердің жалғаулар кестесінің құрылуын келесі суреттен көруге болады.

RecNo	id	okonch	chislo	vremia	lico	ch_r
	Click here to define a filter					
1	1	ю	1	1	1	2
2	2	ешь	1	1	2	2
3	3	ишь	1	1	2	2
4	4	ет	1	1	3	2
5	5	ем	2	1	1	2
6	6	ете	2	1	2	2
7	7	ите	2	1	2	2
8	8	ют	2	1	3	2
9	9	л	1	2	4	2
10	10	ла	1	2	4	2
11	11	ло	1	2	4	2
12	12	ли	2	2	4	2
13	13	ать	5	5	5	2
14	14	ыть	5	5	5	2
15	15	ить	5	5	5	2
16	16	еть	5	5	5	2
17	17	ять	5	5	5	2
18	18	уть	5	5	5	2
19	19	у	1	1	1	2
20	20	лю	1	1	1	2

1 сурет. Орыс тілі етістіктердің аффикстер кестесі.

Жоғары суреттегі етістік аффикстеріне төрт атрибут (число, время, лицо, часть речи) жалғанған, осы атрибуттарға сәйкес машиналық аударманың екінші (аудару) тілінің аффиксі ізделінеді. 1 суретте көрсетілгендей талдау және іздеу кезінде қызылмен ерекшеленген көпмағыналық бейнелеулерге тап боламыз.

Ал аударма процесінде көпмағыналы жағдайда біз бірмағыналы шешімді табуымыз қажет. Сол себептен әр көпмағыналы жағдайға әдейі арналған бірмағыналы шешімдер шығаратын ережелер құрамыз.

2 ОРЫС ЖӘНЕ ҚАЗАҚ ТІЛДЕРІНІҢ СЕПТІЛІК СӘЙКЕСТЕНДІРУ КЕСТЕЛЕРІ

Қазақ тілін орыс тілімен салыстырғанда септіктердің мағынасына қарай салыстыру жүйесін жасадық. Бірақ мұнда өзіндік қиыншылықтар туындады, себебі орыс тілінде алты, ал қазақ тілінде жеті септік бар және олар бір біріне сәйкес емес. Олардың сөз түрлендіруі бір бірінен өзгеше. Қазақ тілінің 7 септік түрі бар және кестеге еңгізу үшін біз оларға реттік нөмірлерді бердік: атау-1, ілік-2, барыс-3, табыс-4, жатыс-5, шығыс-6, көмектес-7. Орыс тілімен қазақ тілін сәйкестендіре отырып септіктің байланысын орнаттық (1 кестеде бейнеленген).

Кесте 1. Екі тіл септіктерінің сәйкестігі.

Орыс тіліндегі септік түрлері	Қазақ тіліндегі септік түрлері
Именительный падеж	Атау септік
Родительный падеж	Ілік септік
Дательный падеж	Барыс септік
Винительный падеж	Табыс септік
Творительный падеж	Көмектес септік
Из (предлог)+ родительный падеж	Шығыс септік
От (предлог)+ родительный падеж	Шығыс септік
На (предлог)+ предложный падеж	Жатыс септік
В (предлог)+ предложный падеж	Жатыс септік

3 ОРЫС ТІЛІНЕН ҚАЗАҚ ТІЛІНЕ МАШИНАЛЫҚ АУДАРМАСЫНЫҢ МОРФОЛОГИЯЛЫҚ ТАЛДАУ МЕН СИНТЕЗІН ШЕШУ

Орыс тілінен қазақ тіліне машиналық аударма бағытында септіктерге қатысты мәселелерді шешу үшін біз **көпмағыналық бейнелеу кестелерін** енгіздік, яғни деректер базасы кестелерден тұрады және әрбір кестеде өзіндік атрибуттары болады. Әр кестеде атрибуттар тізімі екі топқа бөлінеді: ену тобы және шығу тобы. Мысалға орыс тілінен мынадай бір сөз тіркесі келіп түсті делік: «гордиться городом» - қаламен мақтану. Алдымен жалғауды орысша сөз үшін қарастырады, ол орыс тілінің септіктерінің жалғауларынан қарайды. Кестеде 5 суретке сәйкесінше былайша болады:

RecNo	id	okonch	ch_r	padezh	chislo	skl
30	30	я	1	2	1	2
31	31	у	1	3	1	2
32	32	ю	1	3	1	2
33	33	<null>	1	4	1	2
34	34	о	1	4	1	2
35	35	е	1	4	1	2
36	36	ом	1	7	1	2
37	37	ем	1	7	1	2
38	38	ём	1	7	1	2
39	39	е	1	5	1	2
40	40	ы	1	1	2	2
41	41	и	1	1	2	2
42	42	я	1	1	2	2
43	43	ов	1	2	2	2
44	44	ей	1	2	2	2
45	45	ам	1	3	2	2
46	46	ям	1	3	2	2
47	47	ы	1	4	2	2
48	48	ей	1	4	2	2
49	49	я	1	4	2	2
50	50	ы	1	6	1	1
51	51	и	1	6	1	1

Сурет 2. Орыс тіліндегі «творительный» септігін анықтау мысалы

Бұл бейнелеу кестеде ену атрибуттар тобы- «okonch» (орыс тілі жалғауы), ал шығу атрибуттар тобына қалғандары кіреді- ch_r, padezh, chislo, skl. Бұл жерден «город+ом» жалғауын тапты, енді оның атрибуттарын қарастырады, ch_r- сөз табы, okonch- жалғау, padezh- септік. Сол сандар, яғни негізгі сипаттаушы параметрлер бойынша енді қазақ тілінің жалғаулар кестесіне барады(Сурет 3):

RecNo	id	okonch	ch_r	padezh	chislo	dauysty	dauyssyz
44	44	терде	1	5	2	2	3
45	45	нан	1	6	1	1	2
46	46	нен	1	6	1	2	2
47	47	дан	1	6	1	1	1456
48	48	ден	1	6	1	2	1456
49	49	тан	1	6	1	1	3
50	50	тен	1	6	1	2	3
51	61	лардан	1	6	2	1	14
52	62	лерден	1	6	2	2	14
53	63	дәрдан	1	6	2	1	256
54	64	дерден	1	6	2	2	256
55	65	тардан	1	6	2	1	3
56	66	терден	1	6	2	2	3
57	67	мен	1	7	1	1	1245
58	68	бен	1	7	1	2	6
59	69	пен	1	7	1	1	3
60	70	лармен	1	7	2	2	14
61	71	лермен	1	7	2	1	14
62	72	дармен	1	7	2	2	256
63	73	дермен	1	7	2	1	256
64	74	тармен	1	7	2	2	3
65	75	термен	1	7	2	1	3

Сурет 3. Қазақ тілінің көмектес септігіне сәйкестік мысалы

Бұл кестеде шығу атрибуты – ‘okonch’(қазақ тілі жалғаулары), ал ену атрибуттар тобы- қалған атрибуттар. 3 суреттен байқап отырғанымыздай бұл жерден орыс тіліндегі «ом» жалғауының {1 7 1} деген сандары (атрибуттары) қазақ тілінің «мен»-{ 1 7 1} деген жалғауымен сәйкес келіп тұр. Сондықтан осыны алып қазақ тіліне аударылған түбір сөзге жалғайды. Сол уақытта «городом»- «қаламен» деп дұрыс аударманы шығарады.

Келесі кезекте етістік үшін морфологиялық синтезді қарастыралық. Бұл жағдайдағы жалғаудың жалғану заңдылығы да дәл сондай принциппен жұмыс істейді. Мысалы кіріс тіліндегі сөз тіркесін тағы да «гордиться городом» деп алайық. Бұл жағдайда «гордиться» сөзі (етістігі) етістіктің алғашқы формасында (неопределенная или начальная форма глагола) тұр. Ал етістіктің алғашқы формасы қазақ тіліндегі тұйық етістік ұғымымен сай келеді. Ондай сөздерді базаға дәл сол қалпында енгізуді жөн көрдік. Себебі олардың түбірі кейбір жағдайларда өзгеріп отырады, мысалы «гордиться» сөзі «горжусь» деген формада да

бола алады, ал бұл істі қиындата түседі, сол үшін алғашқы формасы базада тұрады. Яғни аударма мынадай түрде болады: «гордиться городом» - қала**мен** мақтану.

Басқаша мысалды алып қарайтын болсақ, «читаешь книгу». Бұл жерде зат есімнің жалғауын қалай анықтайтындығы жайлы жоғарыда сипаттап өттік, енді етістігін қалай анықтайдығын айтайық. Сөз тіркесі келіп түскеннен соң, базадан сөздерді іздейді, сол кезде «читаешь» сөзінің етістіктер кестесінде жатқандықтан етістік екендігін анықтап біледі. Алайда кестеде бұл сөздің тек түбірі ғана болады: «чита». Сол сөздің аудармасының негізі «оқ» екендігін тапқаннан соң, 4-суретке сәйкес

RecNo	id	rus	osnova_rus	kaz
Click here to define a filter				
698	1034	разрядить	разряд	жасандыр
699	1035	наряжаться	наряжа	жасан
700	1036	лежа	леж	жат
701	1037	советовать	совет	кеңес ет
702	1038	читать	чита	оқ
703	1039	помнить	помн	еске ал
704	1040	взять	возьм	ал
705	1041	воевать	вою	соғыс
706	1042	воровать	ворова	ұрла

Сурет 4. Етістіктер сөздігінің кестесі

«ешь» жалғауын жалғаулар кестесінен іздейді, ол 5-суретте келтірілген:

RecNo	id	okonch	chislo	vremia	lico	ch_r
Click here to define a filter						
1	1	ю	1	1	1	2
2	2	ешь	1	1	2	2
3	3	ишь	1	1	2	2
4	4	ет	1	1	3	2
5	5	ем	2	1	1	2
6	6	ете	2	1	2	2
7	7	ите	2	1	2	2

Сурет 5. Орыс тілі етістіктің жалғаулар кестесі

Көріп тұрғанымыздай «ешь» жалғауын тапты, енді оның атрибуттары бойынша, яғни *okonch* (жалғау)-«ешь», *zhak* (жақ)- 2, *shak* (шақ)- 1 және *tur*(түр)- 1 «жекеше» екендігін тауып алды. Сол бойынша *etistik_kaz* кестесінен дәл сондай сандарды (атрибуттарды) іздейді. Ондай атрибуттар саны 6-суретке сәйкес екеу болып тұр:

RecNo	id	okonch	zhak	shak	tur	dauysty	dauyssyz
Click here to define a filter							
1	1	мын	1	1	1	<null>	<null>
2	2	мін	1	1	1	<null>	<null>
3	5	пын	1	1	1	<null>	<null>
4	6	пін	1	1	1	<null>	<null>
5	7	сың	2	1	1	1	4
6	8	сің	2	1	1	2	4
7	9	мыз	1	1	2	<null>	<null>
8	10	міз	1	1	2	<null>	<null>
9	11	быз	1	1	2	<null>	<null>

Сурет 6. Қазақ тілі жалғаулары бойынша екі сәйкестік табылған жағдай

Алайда ол атрибутқа сай келетін жалғау екеу болып тұр. Оның бірі «сың» сингармонизм заңы бойынша буын жуан болған жағдайда, ал екіншісі «сің» буын жіңішке болған жағдайда жалғанады. Соны компьютерге түсінікті ету үшін қосымша тағы бір атрибуттар қосылады.

Бұл жердегі дауысты 1- ол жуан, дауысты 2- жіңішке, ал дауыссыз 4- «р, й, у» әріптерінен соң осындай жалғау жалғанатындығын көрсетуші. Ол әріптер комбинациясы программада енгізілген. Яғни осы ережелер бойынша «ешь» жалғауына қазақша сәйкесі «сың» болады. Сондағы алатын нәтижеміз «читаешь книгу»- «кітап оқып отырсың». Бұл жерде тағы бір ескертетін жайт, қазақ тілінде көмекші етістік деген ұғым бар. Ол негізгі етістікке көсемше жалғауын тіркеген соң қосылады. Ал көсемше жалғаулары программада айтылады. Жуаннан кейін «ып», жіңішкеден кейін «іп».

Біздің деректер базасында, зат есімдерге жалғанатын көптік жалғаулар кестесі екеу, бірі орыс тілінің көптік жалғаулары және екіншісі қазақ тілінің көптік жалғаулары. Біздің мақсатымыз кіріс тілінде келген жалғаудың көптік екендігін анықтап, оны атрибуттар арқылы қазақ тіліндегісімен сәйкестендіру. Мысалы мынадай сөз тіркесі берілсін: «книги лежат на полке». Бұл жердегі «книги» сөзінің көптік жалғау ма я септік жалғауы ма екендігін алдымен программа анықтап алуы керек. Егер, сөз тіркесі не сөйлемдегі етістік көпше түрде тұрса, онда бұл жалғау да сәйкесінше көптік жалғауы болады. Ал біздің жағдайда «лежат» сөзіндегі жалғау көпше түрдегі заттың әрекетін білдіріп тұр. Сол себепті зат есім де көпше түрде. Мұны анықтап алғасын енді енді «книг» сөзін кестелерден жүгірте шығып, іздеп тапқасын аудармасын аламыз. Кейіннен оның жалғауы «и» тағы кестелерден жүгірте отырып ізделінеді, табылғасын оның атрибуттарын қарап 3-суреттегідей, қазақша балама жалғауды атрибуттар сәйкестігі бойынша табамыз:

RecNo	id	okonchanie	ch_r	padezh	chislo
Click here to define a filter					
1		и	1	1	2
2	2	ы	1	1	2
3	3	я	1	1	2
4	4	а	1	1	2

(а)

RecNo	id	okonchanie	ch_r	padezh	chislo	dauysty	dauyssyz
Click here to define a filter							
1	1	лар	1	1	2	1	3
2	2	лер	1	1	2	2	3
3	3	дар	1	1	2	1	2
4	4	дер	1	1	2	2	2
5	5	тар	1	1	2	1	1
6	6	тер	1	1	2	2	1

(ә)

Сурет 7. Орыс және қазақ тіліндегі көптік жалғаулар кестесі.

Енді қазақ тілінен {1 1 2 } атрибуттарымен жалғау іздейміз. Байқап отырғанымыздай бұл жерде барлық жалғау біз іздегендей атрибуттар тұр. Бұл жердегі дауысты 1- дауысты жуан, дауысты 2- дауысты жіңішке, дауыссыз 1- «б,в,г,д» және барлық қатаң дауыссыздар, дауыссыз 2- «м,н,ң» дауыссыздары деп белгіленген

Яғни бізге келіп түскен «книги» сөзін не «кітаплар» немесе «кітапдер» деп емес, «кітаптар» деп дұрыс аударып беруі керек. Осылайша қажетті әрі дұрыс аударма шығарылады: «кітаптар сөреде жатыр».

Осындай салдарда лингвистика заңдарына жүгінеміз. Жалғауды дұрыс тіркеу үшін сингармонизм заңына бағынамыз. Қысқа айтып кетсек: сингармонизм заңы бұл сөздің түбірі қай әріпке аяқталанытына байланысты жалғауға қолданылатын үндестік ережелер жиынтығы. Мұндай ережелерді автоматандыруға қиындық соқпайды, себебі қазақ тілінің дыбыс үндестігі заңы тұрақты және барлық түркітілдес тілдерге тән.

Жоғарыда көріп тұрғандай аффикстердің жалғануы морфологиялық қасиеттерінен қана емес, сонымен қатар тіркеленетін сөздің түбіріне де байланысты. Бірақ кейде мұндай ақпарат жеткіліксіз болып қалады, себебі аффикстердің жалғануы мәтіндегі басқа тәуелді сөздерге де байланысты. Көрсетілген мысалда сөздің жалғауын дұрыс анықтау үшін «лежат» етістіктің көмегіне жүгіндік. Сонымен қатар бізге тек қана іс-әрекет байланысы ғана емес

тұлғасын (субъект \ объект) да иелендіру қажет. Мысалы ретінде жіктік және тәуелділік жалғауларды қарастыруға болады.

Орыс тілінде жіктік жалғау болмайды, тек қазақ тіліне аударма кезінде ғана жаққа байланысты жалғау қосылады, мысалы «я студент»- мен студентпін сол сияқты 2-3 жақтар үшін де түрлі формада болады. Оны анықтау үшін деректер базасындағы индекстік файл кестесін пайдаланамыз. Келген есімдіктің атрибуттарын тауып аламыз да 8-суретте келтірілгендей, сол бойынша сәйкес келетін жалғауды табамыз:

RecNo	id	rus	kaz	chislo	lico	padezh
Click here to define a filter						
1	1	я	мен	1	1	1
2	2	ты	сен	1	2	1
3	3	он	ол	1	3	1
4	4	она	ол	1	3	1
5	5	оно	ол	1	3	1
6	6	мы	біз	2	1	1
7	7	вы	сендер (сіздер)	2	2	1
8	8	они	олар	2	3	1

Сурет 8. Жіктік жалғауы және оның атрибуттары

«я» есімдігінің атрибуттары {1 1 1} екендігін анықтағасын енді кестесіне барып {1 1 1} іздейміз:

RecNo	id	zhalgau	chislo	lico	padezh	dauysty	dauyssyz
Click here to define a filter							
1	1	мын	1	1	1	1	1
2	2	мін	1	1	1	2	1
3	3	бын	1	1	1		<null>
4	4	бін	1	1	1		<null>
5	5	пын	1	1	1	1	3
6	6	пін	1	1	1	2	3
7	7	сың	1	2	1	1	123456
8	8	сің	1	2	1	2	123456

Сурет 9. Қазақ тіліндегі бірнеше сәйкес жіктік жалғаулары табылған жағдай

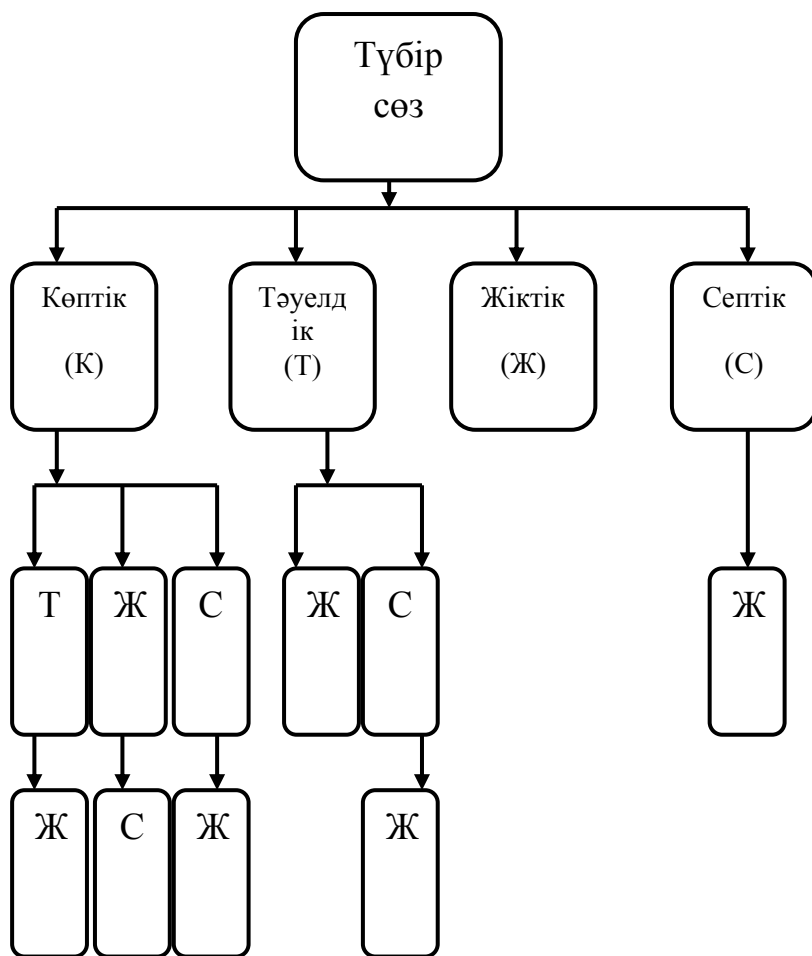
Ол атрибут алғашқы алты жалғаудың көпмағыналы бейнелеу болып, барлығында бірдей екендігі 9-суретте көрсетілген. Оларды ажырату мақсатында жоғарыда айтылып кеткен сингорманизм заңының ережелеріне сүйеніп қосымша дауысты дауыссыз атрибуттары (бағаналары) қосылды. Сол бойынша қажеттісі таңдалып алынады, біздің жағдайда ол «пін», себебі «студент» сөзі қатаң дауыссызға аяқталып тұр. Яғни алатын аудармамыз «я студент»- «мен студентпін»

Тәуелдік жалғау, әдетте, бір заттың басқа бір затқа тәуелді екенін білдіретін қосымша. Негізінде зат есімге тән қосымша бола тұрса да, зат есім қызметін атқаратын, демек, субстантивтенетін (заттанатын) сөздердің барлығына да жалғана береді. Бұл қосымшалар жалғанған сөздер, әдетте, өздерінен бұрын ілік септік жалғауда тұрып тіркесетін жіктеу есімдіктермен тікелей байланысты болады. Сол себептен тәуелдеулі сөздің жақ жалғаулары да жіктеу есімдіктерінің жақтарына сәйкес келіп отырады. Мысалы: менің қаламым; сенің қаламың; сіздің қаламыңыз, оның қаламы.Осылайша сәйкес әрі қажетті жалғау таңдалып алынып, қазақшаға аударылған түбір сөзге жалғанады. Біздің жағдайда «моя парта»- «менің партам»

Жалпы келгенде машиналық аудармада аффикстер сөз құру және түрлендіру ролін атқарып сөз арасындағы байланысты орнататын құралдардың бірі болып табылады. Осындай маңызды морфологиялық процесске көпмағыналы бейнелеу кесте әдісі өте қолайлы болды.

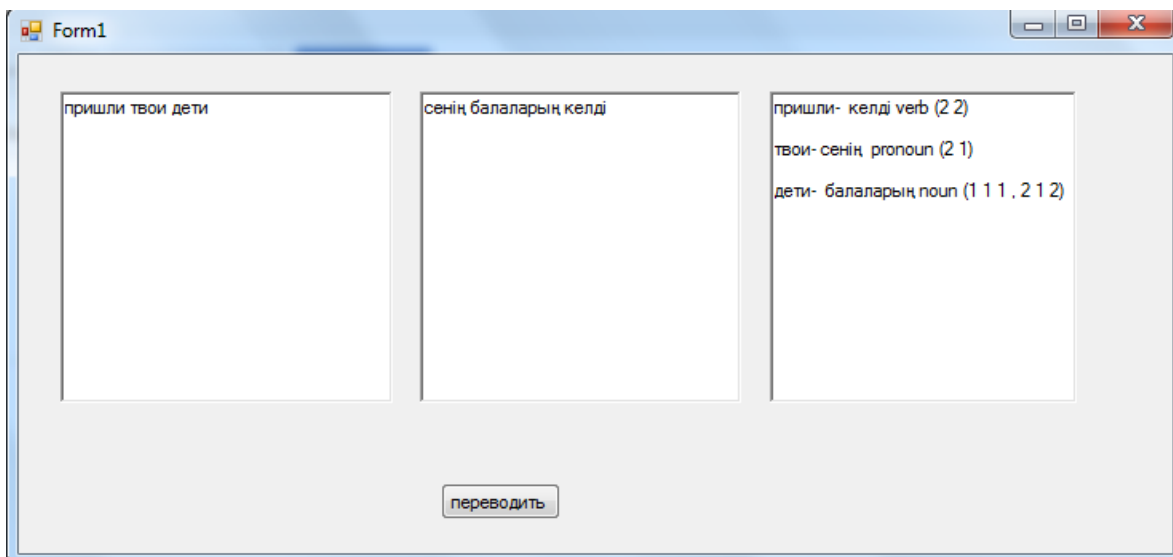
4 КӨПМАҒЫНАЛЫҚ БЕЙНЕЛЕУ КЕСТЕЛЕР АРАСЫНДАҒЫ БАЙЛАНЫС

Көпмағыналы бейнелеу кестелердің құрылуы мен қолдануы жоғары тарауларда көрсетілген. Осы кестелер бойынша алынған мәліметтерді машиналық аудармада шығыс тіліне дұрыс генерациялау білуі қажет. Морфологиялық талдау және синтез кезінде сөзге осындай көпмағыналы бейнелеу кестелер бірнешеу табылуы мүмкін. Себебі орыс тілінің аффикстері көп емес боландықтан сөздің лексикалық та, грамматикалық та мағынасы біріктірілген. Ал қазақ тілінің сөздер құрылымында екі немесе одан да көп аффикстер жалғануы мүмкін. және қазақ аффикстерінің өзіндік жалғану заңдылығы бар. 10-суретте келтірілген диаграммада қазақ тіліндегі жалғаулардың түрі және бір біріне тіркесуінің мүмкін болатын барлық жағдайлары келтірілген.



Сурет 10. Қазақ тіліндегі жалғаулар классификациясы мен комбинациясы.

Бұл өте қажетті талдаулардың бірі, себебі сөйлемді я тіпті сөзді біз белгілі бір заңдылықтарға сүйене отырып қана құрай аламыз, мысалы, септік жалғаудан кейін көптік жалғауды тіркей алмаймыз. Септік жалғаудан кейін тек қана жіктік жалғауы қойылады т.б. Мысалы: «пришли твои дети» мәтінін қазақ тіліне аударып, сөздердің жалғауларына талдау жасайық.



Сурет 11. Орыс тілінен қазақ тіліне машиналық аударманың мысалы

ҚОРЫТЫНДЫ

Біз орыс тілінен қазақ тіліне машиналық аудармадағы морфологиялық талдау және синтез мәселелерін қарастырдық. Сонымен, көпмағыналық бейнелеу машиналық аудармаға көпмағыналық мәселелер туғызады, оларды шешу жолдары жоғарыда көрсетілгендей қосымша атрибуттарды еңгізу, немесе қосымша осы мәселелерді шешетін процедураларды құруды қажет етеді. Морфологиялық талдау мен синтез мәселелеріне тоқтатылып, сәйкес орыс және қазақ тілдерінің грамматикалық қасиеттеріне сүйеніп көпмағыналы бейнелеу кестелер және ережелер жиынтығы құрылған. Екі тілді сәйкестендіріп талдау жасадық. Алынған нәтижелер машиналық аударманың сапасын көтереді. Қазақ тілдегі жалғаулардың барлық мүмкін болатын комбинациялар зертелінген. Осы көпмағыналы бейнелеу кестелер әдісі арқылы грамматикалық заңдылықтарды пайдала отырып, орыс тілінен қазақ тіліне машиналық аудармадағы морфологиялық талдауы мен синтезі жоғары сападағы нәтижелерді көрсетті. Аудармашыдағы деректер қорындағы элементерді іздеу жылдамдығын және таңдау сапасын жоғарлатады. Практикалық нәтижеде алынған зерттеулер Microsoft Visual Studio10 бағдарламасының C# тілінде 10 000 дана сөзі бар орыс-қазақ аудармашыны жасау барысында қолданылған.

ӘДЕБИЕТ

1. Баскаков Н. А. Хасенова А.К. Исенғалиева И.А. Кордабаев Т.Р. Сопоставительная грамматика русского и казахского языков. Морфология. Изд-во "Наука",1966.
2. Введение в теорию многозначных отображений. Составитель Гельман Б.Д., Воронеж, 2003.
3. Tomasz Kaczynski, Multivalued Maps As a Tool in Modeling and Rigorous Numerics. Departement de mathematiques,Universite de Sherbrooke, 2008.
4. Teruko Mitamura, Eric H. Nyberg, Hierarchical lexical structure and interpretive mapping in machine translation, Proceedings of International Conference COLING- 1992 Nantes, 1254-1258 pp.
5. Dilek Zeynep Hakkani, Gökhan Tür, Kemal Oflazer, Teruko Mitamura, and Eric H. Nyberg, An English-to-Turkish Interlingual MT System, Proceedings of International Conference AMTA-1998, pp. 83-94.