

УДК 19.711.3, 510.5, 004.43

ЖУМАНОВ Ж.М.**МАШИННЫЙ ПЕРЕВОД С АНГЛИЙСКОГО ЯЗЫКА НА КАЗАХСКИЙ ЯЗЫК С ИСПОЛЬЗОВАНИЕМ СЕМАНТИЧЕСКИХ СИТУАЦИЙ**

(Казахский национальный университет им. аль-Фараби, г. Алматы, Республика Казахстан)

Текст на любом естественном языке может быть представлен как набор предложений, связанных между собой и несущий определенный смысл. Каждое предложение, в свою очередь, состоит из более мелких смысловых конструкций — словосочетаний и слов. В каждой из этих конструкций можно выделить 2 составляющие: смысл и форму.

Рассмотрим предложение «I have to take my books» («Я должен взять мои книги») можно разбить на 4 части: «I», «have to take», «my», «books». Слова «I» и «my» стоят в словарной форме и могут быть переведены напрямую без затруднений. Сочетание «have to take» - более сложный случай. Все выражение означает что субъект должен что-то выполнить. На это указывает используемая в выражении форма - «have to». Что именно он должен выполнить указывает смысловая часть - «take» («взять»). Слово «books» - представляет собой множественное число слова «book» («книга»). На это указывает использование окончания «-s», которое и является в данном случае признаком формы. Смысл этой конструкции передает само слово.

Смысловые конструкции, в отличие от слова в словарной форме, переводятся в 2 этапа. Сначала необходимо перевести смысловую часть, затем изменить форму перевода так, чтобы она соответствовала той форме которая использована в исходном предложении.

Подобные смысловые конструкции предлагается называть «семантическими ситуациями» (для краткости — просто «ситуациями»). Так как их использование зависит от ситуации в которой употребляется переводимое предложение. Таким образом получаем следующее: Предложения состоят из семантических ситуаций. Ситуациями могут являться: словарные слова (самый элементарный случай), грамматические формы слов (падежные формы, различные числа, лица и т.п.), устойчивые сочетания слов.

Формально текст на естественном языке можно описать следующим образом:

<текст> ::= <предложение> | <текст><предложение>

<предложение> ::= <сочетание слов> | <предложение><сочетание слов>

<сочетание слов> ::= <грамматическая форма слова> | <сочетание слов><грамматическая форма слова>

<сочетание слов> ::= <слово> | <сочетание слов><слово>

<грамматическая форма слова> ::= <слово><грамматические признаки>

<грамматическая форма слова> ::= <грамматические признаки><слово>

<грамматические признаки> ::= <аффикс> | <предлог> | <...>

<слово> ::= <существительное> | <прилагательное> | <...>

А семантические ситуации определяются в этом случае как:

$\langle \text{предложение} \rangle ::= \langle \text{семантическая ситуация} \rangle \mid \langle \text{предложение} \rangle \langle \text{семантическая ситуация} \rangle$
 $\langle \text{семантическая ситуация} \rangle ::= \langle \text{слово} \rangle \mid \langle \text{грамматическая форма слова} \rangle \mid \langle \text{сочетание слов} \rangle$

Перевод текста с использованием данной модели сводится к переводу предложений, из которых он состоит. А перевод предложений — к переводу составляющих его семантических ситуаций.

Алгоритм перевода предложения английского языка на казахский язык с использованием семантических ситуаций представлен на рисунке 1. Поиск семантических ситуаций в предложении осуществляется слева направо.

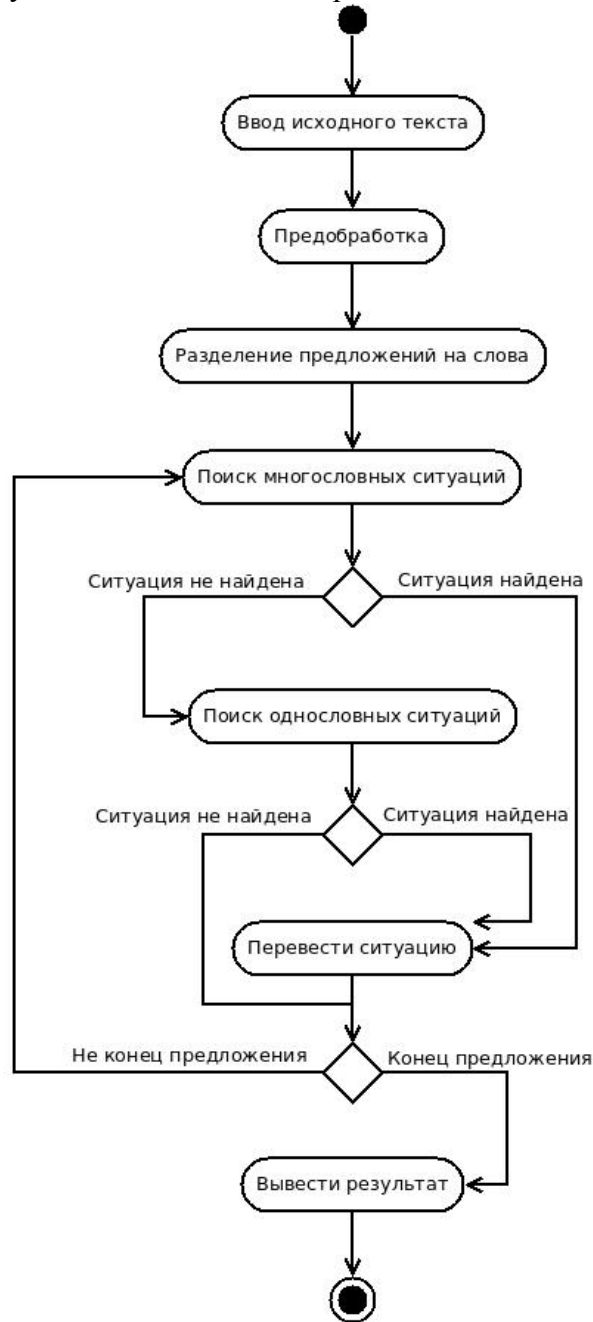


Рисунок 1 — Алгоритм перевода с использованием семантических ситуаций

Сначала ищутся ситуации, состоящие из двух и более слов. В случае если таковых нет, тот же самый сегмент предложения проверяется на наличие более коротких ситуаций. Когда поиск доходит до однословных ситуаций, возможно 3 варианта: находится ситуация — грамматическая форма некоего известного слова, находится словарная форма известного слова, находится неизвестное слово (слово отсутствующее в словаре). В любом случае поиск ситуаций и перевод будет продолжен. Неизвестное слово будет выведено в результате перевода в своем исходном виде.

Программа перевода, работающая по указанному алгоритму в процессе своей работы проходит через последовательность состояний представленных на рисунке 2. Как можно видеть, эта последовательность повторяет алгоритм перевода. Вся последовательность переходов между состояниями повторяется для каждого переводимого предложения.



Рисунок 2 — Диаграмма состояний программы перевода

УДК 519.768.4

КӘРІБАЕВА А.С., ӘМІРОВА Д.Т., ТУКЕЕВ У.А.

**АПЕРТИУМ ПЛАТФОРМАСЫНДАҒЫ ҚАЗАҚ – АҒЫЛШЫН ЖӘНЕ
АҒЫЛШЫН – ҚАЗАҚ МАШИНАЛЫҚ АУДАРМАСЫНДА ЛЕКСИКАЛЫҚ ТАҢДАУ
МӘСЕЛЕСІН ШЕШУ**

(Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан)

1. Кіріспе

Соңғы уақытта халықаралық алаңда Қазақстан Республикасының рөлі өсіп келеді, бұл өзінің кезегінде әлемдік қоғамдастықтың біздің елімізге деген қызығушылығын арттырады. Бүгінгі күні ағылшын тілі халықаралық тіл болып табылады. Қазақ тілі – Қазақстан Республикасының мемлекеттік тілі. Аудармашылардың жұмыс көлемі жылдан жылға артып келеді. Сәйкесінше, ағылшын тілінен қазақ тіліне автоматтандырылған аударманың құралдарын жасау маңызды міндеттердің бірі болып табылады.

Лексикалық көп мәнділігін шешу – бұл белгілі мәнмәтінде көп мағыналы сөздің мағынасын табу процесі болып табылады. Адам үшін мәнмәтінде сөздердің дұрыс мағынасын табу қиынға түспейді, бірақ машина үшін бұл оңай емес. Себебі, адамның сөздердің мағынасын дұрыс түсіну және таңдау қабілетін жасау алгоритмдерін құру өте қиын есеп.

Лексикалық көп мәнділігін шешу ұзақ тарихына қарамастан әліде дамып келе жатқан білім саласы болып табылады. Бірақ бұл сұрақтың шешу қажеттілігінің үлкен маңызы бар, себебі кеңінен қолданылады (ақпараттық ізденіс, ақпаратты шығару, машиналық аударма, контент-анализ, мәтіндерді өңдеу, сөйлеуді өңдеу).

Лексикалық таңдау мәселесін Апертиум ашық кодты/тегін платформасында лексикалық таңдау ережелерін жазу арқылы шешіп жатырмыз(Forcada et al. 2011, <http://www.apertium.org>)[0.

Апертиум – бұл бастапқы кодтармен берілген ашық машиналық аудармашы жүйесі.

Апертиум машиналық аударма жүйесі бірнеше модульдерден құралады, солардың бірі – лексикалық модуль [0.

2. Apertium платформасындағы лексикалық таңдау ережелері

Лексикалық таңдау Апертиум платформасындағы басқа модульдар сияқты аударма жүйесін ақпаратпен қамтамасыз етеді. Осы ережелер формализі ережелердің мәнмәтініне негізделеді. Ал, ережелер мәнмәтіні келесідей белгілердің тізбектілігінен тұруы керек:

Берілген тілдің жеке лексикалық формасынан;

Мақсат тілдің жеке лексикалық формасынан;

Келесі бір операциялардың бірінен:

SELECT – мақсат тілдің лексикалық формасына сәйкес аудармасын таңдайды және ол таңдалмаған басқа аудармаларды жояды;

REMOVE – берілген лексикалық форма шаблонын қанағаттандыратын мақсат тілдің аудармасын жояды;

Лексикалық таңдау ережелерін жазуда қолданылатын шаблондар:

<rule> - ереженің басы;

<match lemma="берілген_тілдегі_сөз"> – анықталатын сөз;

tags="сөз_табы" – анықталатын сөздің жататын сөз табының тәгі, мысалы, зат есім - "n", сын есім - "adj", т.с.с.;

<select lemma="таңдалатын_сөз" – көпмағыналы «анықталатын сөздің» мәтін мағынасына сәйкес бір аудармасын таңдау;

tags="сөз_табы" – таңдалатын сөздің қай сөз табына жататынын көрсететін тәг;
</match>, </rule> - сәйкес тәгтердің жабылуы.

Лексикалық таңдау – сөз көп мәнді сөздің аудармасын мәтін мәнмәтініне байланысты бірін таңдау.

Лексикалық таңдау – Апертиум платформасындағы модульдердің бірі. Бұл модуль сөз таптарын таңдаудан кейін жұмыс жасайтын модульдердің бірі.

Лексикалық таңдаудың келесі ұғымдары бар:

Аударылатын сөз;

Ізделінетін сөз, яғни алатын аударма нұсқасы;

Мәнмәтін. Бұл белгілер мен орналасқан орынның жиыны;

Операциялар. Бұл аударылатын сөзді таңдаудан немесе жоюдан тұрады.

Лексикалық таңдау ережелері қолмен жазылады. Ережелерді жазу алдында алдымен файлы құрылады. Оның аты келесідей болады: apertium-eng-kaz.kaz-eng.lrx (қазақ - ағылшын үшін), apertium-eng-kaz.eng- kaz.lrx(ағылшын - қазақ үшін).

Лексикалық таңдау ережелерін жазар алдында, алдымен қазақ тілінен ағылшын тіліне аударғанда бірнеше аудармасы бар қазақ тіліндегі сөзді таңдап алуымыз керек. Мысалы, “жер” деген сөздің алты аудармасы болады:

Жер – ground

Жер – place

Жер – land

Жер – earth

Жер – parcel

Жер – spot

Жалпы осы берілген сөздерді сөздікке енгізгенде ең жиі кездесетіні бірінші енгізіледі. Осы берілген сөздер әртүлі тіркеспен келгенде, әртүрлі болып аударылады. Егер “қазақ жері” деген тіркесті аударғанда “Kazakh land”, “жер планетасы” тіркесінің аудармасы “earth

planet”, ал “алыс жер” тіркесінің аудармасы “distant place” болады. Осы сөзге байланысты ереже келесідей болады:

```
<rules>
  <rule>
    <or>
      <match lemma="қай" tags="det.*"/>
      <match lemma="барлық" tags="det.*"/>
    </or>
  <match lemma="жер"><select lemma="place"/></match>
</rule>
<rule>
  <match lemma="қазақ" tags="n.attr.*"/>
  <match lemma="жер" tags="n.px3sp.nom.*">
  <select lemma="land" tags="n.*"/></match>
</rule>
<rule>
  <match lemma="жер" tags="n.attr.*"/>
  <select lemma="earth" tags="n.attr"/></match>
  <match lemma="планета" tags="n.px3sp.*.*">
</rule>
```

Қазақ тілінен ағылшын тіліне аударғанда кейбір сөздердің бірнеше аудармасы болатынын білдік. Енді тағы бір мысалды қарастырайық. Қазақ тілінде тек категориясы болмағандықтан, “ол” есімдігінің тек бір ғана аудармасы болады. Ал, қазақ тілінен ағылшын тіліне аударғанда үш аудармасы болады. Жалпы жағдайда, осы есімдіктің үнсіздік бойынша аудармасы “he” болады. Ал, егер нақты “Ол әдемі қыз” деген сөйлемде, “she” есімдігі болуы керек. Осы берілген сөйлемге лексикалық таңдау ережесін жазабаса, онда жүйе келесідей аударманы береді:

He is beautiful girl(бұл аударма дұрыс емес)

Аударғанда “She is beautiful girl” болуы керек. Осындай аударма болуы үшін келесідей лексикалық таңдау ережесін жазамыз:

```
<rule>
  <match lemma= “Ол” tags= “prn.pers.p3.sg.nom”>
  <select lemma= “she” tags= “prn.subj.p3.f.*”/> </match>
  <match lemma= “әдемі” tags= “adj”>
  <select lemma= “beautiful” tags= “adj”/></match>
  <match lemma= “қыз” tags= “n.*”/>
</rule>
```

Осы ережені жазғаннан кейін бұл сөйлем аудармасы дұрыс болады.

Ал, ағылшын тіліндегі «beautiful» сөзінің қазақ тілінде екі аудармасы болады: «әдемі» және «көркем». Осы екі аударма ағылшын-қазақ екітілді сөздігінде - “apertium-eng-kaz.eng-kaz.dix” көрсетіледі. Мәтіннің құрылымына сәйкес, осы екі аударманың лексикалық ережемен біреуі таңдалады. Мысалы, мәтінде келесідей құрылым кездесе, «beautiful girl plays in garden», онда лексикалық ереже келесі түрде жазылады: егер «beautiful» сын еісімінен кейін «girl» зат есімі келсе, онда «beautiful» сөзі «әдемі» деп аударылады. Басқа жағдайда, егер «beautiful place» тіркесінде сын есім «көркем» деп аударылады. Үндеместік бойынша таңдалатын аударма үшін ережені жазуға болады[**Ошибка! Источник ссылки не найден.**

4. Қорытынды

Қазіргі кезде Ағылшын-Қазақ машиналық аудармашысындағы лексикалық ережелер саны оннан, ал Қазақ – Ағылшын аудармашысында он бестен асады. Жалпы лексикалық

таңдау ережелері мақсат тілдегі көпмағыналылықты жоюға мүмкіндік береді. Болашақта жүйені ары қарай дамытамыз деп жоримыз. Жүйе аударма сапасы жоғары болуы үшін әлі де ережелер жазу керек.

Пайдаланылған әдебиеттер тізімі:

1. Forcada, M.L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J.A. Sánchez-Martínez, F., Ramírez-Sánchez, G., Tyers, F.M. 2011. "Apertium: a free/open-source platform for rule-based machine translation". Machine Translation 25(2)127-144.

2. Lexical selection. – URL: http://wiki.apertium.org/wiki/Lexical_selection(13.11.2014)

3. Сундетова А. М., Апертиум платформасындағы Ағылшын-қазақ машиналық аударма лексикалық модулі. Международная научная конференция студентов и молодых ученых «Фараби әлемі». – Алматы: «Қазақ университеті», 2014. – С. 145.

ӘОЖ 81 322

МУКАНОВА А.С.

ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕЛЕРДЕГІ БІЛІМДЕРДІ БЕЙНЕЛЕУ МОДЕЛЬДЕРІН ТАЛДАУ

(Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан)

Қазіргі кезде күрделі есептерді шешу үшін қолданылатын программалық жүйелердің саны артуда. Олар күрделі процестерді басқару кезінде шешім қабылдау жағдайларын бағалауда, техникалық диагностика жасауда, сұрақ- жауап беру кезінде қолданылады. Бұл жүйелер өзіндік оқытылады, өз шешімдерін негіздей алады. Сол себепті де оларды интеллектуалды жүйелер деп атайды. Оларға эксперттік жүйелерді, шешім қабылдау жүйелерін, мәтіндер мен сөйлеулерді тану және т.б. жүйелерді жатқызуға болады.

Интеллектуалды жүйелерді құрудың негізгі мәселелерінің бірі білімдерді бейнелеу моделін [1] таңдау болып табылады. Осы білімдерді бейнелеу моделі жүйенің қасиеттері мен мүмкіншіліктерін, архитектурасын, сонымен қатар, білімдерді алудың әдістерін анықтауға негіз болады. Қазіргі кезде білімдерді бейнелеу модельдерінің бірнеше түрі бар. Олар: фактілер мен ережелер негізінде бейнелеу, предикаттарды есептеу, нейрондық желілер, семантикалық желілер, фреймдер. Осы модельдердің әрқайсысы қандай да бір интеллектуалды жүйені жасауға мүмкіндік береді. Сол себепті де интеллектуалды жүйені жасамас бұрын әрқашан да осы модельдердің қайсысын алу керек деген есеп тұрады.

Білімдерді бейнелеу модельдерін келесі критерийлер бойынша бағалауға болады: білім элементтерінің күрделілік деңгейі, білімді бейнелеудің универсалдылығы, яғни, білімдерді әртүрлі пәндік аймақтар арқылы бейнелей білу мүмкіндігінің болуы, қолдану кезінде білімдерді бейнелеудің көрнекілігі мен табиғиылығы, модельдің жаңа білімдер алуға қабілеттілігі және оларды оқыта білуі, модельдің жад көлемін алатын өлшемі бойынша, берілген модель арқылы жүйені құрудың ыңғайлылығы.

Қазіргі кездегі білімдерді бейнелеу модельдерін жоғарыда көрсетілген критерийлер бойынша талдап көрейік.

Фактілер мен ережелер негізінде білімдерді бейнелеу моделі (продукционды модель) ЕГЕР (шарт)- ОНДА (әрекет) өрнегі арқылы жүзеге асырылады. Егер есепте көрсетіліп тұрған жағдай ЕГЕР ережесіне бағынатын болса, ОНДА бөлігі арқылы анықталатын қызмет жүзеге асады [2]. ЕГЕР ережесінің шартын қанағаттандыратын фактілерді іздеу бізге шығыстар тізбегі - шешімдер ағашын береді. Бұл білімдерді бейнелеу моделінің кемшілігі шығыстар тізбегін жасауға кететін уақыттың көптігінде. Модель қарапайым, біртепті есептерді шешуге арналған, сол себепті де көптепті күрделі есептерді шешу кезінде оның