

А.Ә. Шәріпбай¹, Б.Ш. Разахова¹, А.С. Муканова¹, Л. Жеткенбай¹

¹Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Нұр-Сұлтан қ., Қазақстан

ТҮРКІ ТІЛДЕРІ ГРАММАТИКАСЫНЫҢ ТЕЗАУРУСЫНА ҚОЙЫЛАТЫН ТАЛАПТАР

Аңдатпа

Бұл жұмыста қазақ тілі грамматикасы бойынша тезаурусқа қойылатын талаптар қарастырылады. Дескрипторлар сияқты элементтер, сонымен қатар, аскрипторлар немесе дескрипторлар емес деп аталатын тезаурусқа қосылған синонимдік қатардың терминдері қарастырылады және олар лайықты дескрипторларды табуға көмектесетін қосымша элементтер (мәтіндік кірістер) ретінде пайдаланылады. Дескрипторларға қойылатын талаптар келтірілген. Егер дескриптор сөз тіркесі арқылы құралса және онда тірек сөзі ретінде зат есім қолданылса, онда келесі шарттардың бірі орындалуы тиіс. Және оған мысалдар берілген. Бірмәнді және анық дескрипторды таңдау мүмкін емес жағдайлар болады, мұнда дескриптор ретінде таңдалған термин релятор мен пікірлермен қамтылады. Дескрипторлар арасында ассоциативті қатынасты орнатудың негізгі мақсаттары, қатынастың қандай типтері ассоциативті семантикалық қатынасқа жататыны сипатталған және әрбір қатынас типіне анықтама берілген.

Түйін сөздер: Тезаурус, лексикалық бірлік, ассоциативті қатынастар, семантикалық қатынастар, компьютерлік лингвистика.

Аннотация

А.Ә. Шәріпбай¹, Б.Ш. Разахова¹, А.С. Муканова¹, Л. Жеткенбай¹

¹Евразийский национальный университет им. Л.Н.Гумилева, Нур-Султан, Казахстан

ТРЕБОВАНИЯ К ТЕЗАУРУСУ ГРАММАТИКИ ТЮРКСКИХ ЯЗЫКОВ

В данной работе рассматриваются требования к тезаурусам по грамматике тюркязычных языков на примере казахского языка. Рассматриваются такие элементы, как дескрипторы, а также термины из синонимического ряда, который включен в тезаурус, называются аскрипторами или недескрипторами и используются как вспомогательные элементы (текстовые входы), помогающие найти подходящие дескрипторы. Также приведены требования к дескрипторам. Если дескриптор образован словосочетанием и в нем опорным словом будет существительное, то должно выполняться одно из следующих условий. И к ним приведены примеры. Бывает случаи, когда не удастся подобрать однозначный и ясный дескриптор, то термин, который взят в качестве дескриптора, снабжается релятором или комментарием. Описаны основные цели установления ассоциативных отношений между дескрипторами, какие типы отношения относятся к ассоциативным семантическим отношениям и даны определения к каждому типу отношений.

Ключевые слова: Тезаурус, лексическая единица, ассоциативные отношения, семантические отношения, компьютерная лингвистика.

Abstract

THESAURUS REQUIREMENTS GRAMMARS OF TURKIC LANGUAGES

Sharipbay A.¹, Razakhova B.¹, Mukanova A.¹, Zhetkenbay L.¹

¹L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan

This paper describes the requirements for thesaurus grammar of Kazakh language. Elements such as descriptors are considered, as well as terms from the synonymic series that is included in the thesaurus, called ascriptors or non-descriptors, and are used as auxiliary elements (text inputs) to help find suitable descriptors. The requirements for descriptors are also given. If the descriptor is formed by a phrase and the reference word is a noun, then one of the following conditions must be met. And examples are given for them. There are cases when it is not possible to find an unambiguous and clear descriptor, then the term that is taken as a descriptor is provided with a relator or comment. The semantic relationship between descriptors is described in detail. The main goals of establishing associative relations between descriptors are described. Definitions are given for each type of relationship.

Keywords: Thesaurus, lexical unit, associative relations, semantic relations, computational linguistics.

1. Кіріспе

Ақпараттық кеңістіктің жаһандануына және әлеуметтік желілердің кеңінен таралуына, сондай-ақ, қоғамның интеллектуалды қызметін цифрландыруды жүргізуге байланысты үлкен көлемді табиғи тілдердегі мәтіндік ақпараттық көздерді пәндік аймақтың тезаурустарын қолдана отырып өңдеу қажеттілігі күрт өсті.

Тезаурус – бұл сөздер мен сөз тіркестері мағыналарының жақындығына байланысты түсінік, концепт немесе дескрипторлар деп аталатын бірліктерге топтастырылған сөздік және осы түсініктер (концептер, дескрипторлар) арасындағы семантикалық қатынастар айқын (қатынастар, иерархия түрінде) көрсетіледі [1].

Қазіргі кезде тезаурустарды көрсетудің форматын реттейтін бірқатар стандарттар бар, мысалы ISO 2788 Құжаттама. Біртілді тезаурустарды құру және дамыту бойынша нұсқаулық ISO 5964 Құжаттама – Көптілді тезаурустарды құру және дамыту бойынша нұсқаулық, ГОСТ 7.25-2001 «Біртiлді ақпараттық-іздеу тезаурус. Құрастыру ережесі, құрылымы, құрамы және көрсетілу түрі» және ГОСТ 7.24-2007. Ақпарат, кітапхана және баспа ісі бойынша стандарттар жүйесі. Көптілді ақпараттық-іздеу тезаурус. Құрамы, құрылымы және құрастырылуындағы негізгі талаптар), ANSI/ NISO Z39.19-2005 Құрастыру, формат және біртiлді бақыланатын сөздіктермен басқару бойынша нұсқаулық [2-6].

Информатикада тезаурус қолданылуының басты қажеттілігі – бұл ақпараттық іздеу.

Компьютерлік лингвистика және мәтіндерді автоматты өңдеу саласындағы аса танымал лексикалық ресурстардың бірі WordNet компьютерлік тезаурус болып табылады [1].

Кез-келген пәндік аймақта негізгі ақпараттық бірлік синонимдер деп саналатын терминнің бірнеше нұсқасымен берілетін түсінік болып табылады. Осындай синонимдердің арасынан бір термин таңдалып алынады, ол тезаурус шеңберіндегі түсінікке сілтеменің негізгі тәсілі ретінде қарастырылады да *дескриптор* деп аталады. Синонимдік қатардағы тезаурусқа қосылған басқа терминдер *аскрипторлар* немесе *дескрипторлар емес* деп аталады және сәйкес дескрипторды табудағы қосымша элементтер (мәтіндік кірістер) ретінде қолданылады.

Тезаурустар түсіндірме сөздіктерге ұқсас, бірақ олардың өзара айырмашылықтары бар. Егер түсіндірме сөздіктер берілген терминмен көрсетілген (осы терминнің мағынасын ашу) түсінікті іздеуге қызмет етсе, тезаурустар берілген түсініктің көрсетілуіне сәйкес келетін терминді іздеуге мүмкіндік береді, яғни, түсіндірме сөздікте термин бойынша оның мағынасы ізделеді, ал тезауруста берілген мағына бойынша сол мағынаны білдіретін термин табылады. Сондықтан тезаурустың түсіндірме сөздікке қарағанда терминнің мағынасын тек оның анықтамасының көмегімен емес, сонымен қатар оны басқа терминдермен немесе оның топтарымен олардың арасында семантикалық қатынастар орнату негізінде салыстыруға мүмкіндік береді. Демек, тезаурус негізгі ұғымдардан тұратын терминдер жинағы мен олардың арасындағы берілген семантикалық қатынастар жиынтығымен бірлестіктері болып табылады. Тезауруста *терминдер анықтамасының* болуы оны тек қолмен немесе автоматтандырылған индекстеуге арналған құрал ретінде ғана емес, сонымен қатар пәндік аймақ туралы білімнің жүйеленген көзі ретінде пайдалануға мүмкіндік береді. Қазақ тілі грамматикасынан [7] белгілі негізгі ұғымдарды пайдаланып тезаурус құру үшін қойылатын талаптар мен қатынастарға тоқталайық.

2. Дескрипторларға қойылатын талап

Көпмағыналы табиға тілдің сөзіне бірнеше дескриптор, ал бірнеше *синонимдік* сөздер мен сөйлемшелерге – бір дескриптор сәйкес келеді. Дескрипторлар келесі іздеулер үшін құжаттар мен ақпараттық сұраныстарды координатты индекстеу үшін қолданылады.

Тезаурустағы барлық дескрипторлар келесі шарттарды қанағаттандыруы керек:

- 1) Бірмәнділік – дескриптордың мағынасы бірмәнді болуы керек;
- 2) Түсініктілік – дескриптордың мағынасы тезаурустың пәндік аймағы шеңберінде қолданушыға түсінікті болуы керек;
- 3) Толықтық – дескрипторлар тезаурустың барлық түсініктері мен олардың пәндік аймақтағы байланысын көрсету керек;
- 4) Эквиваленттілік – мағынасы бойынша жақын, бірақ әр түрлі түсінікті білдіретін дескрипторлар эквиваленттілік класстарына біріктіріліп, олардың индекстелуінің субъективтілігін азайту үшін және көпмағыналылықты жою үшін бір дескрипторға келтірілу керек.

Егер дескриптор сөз тіркесінен жасалған болса және ондағы тірек сөз зат есім болса, онда келесі шарттардың бірі орындалуы қажет:

- 1) Дескриптордың мағынасы оның компоненттерінің мағынасынан алынбайды. Мысалы: *сөйлем мүшелері*;
- 2) Дескрипторда компоненттерінің кем дегенде біреуі басқа тіркестер құрамында қолданылмайды немесе үнемі басқа мағынада қолданылады. Мысалы: *зат есім*;
- 3) Дескриптордың толық синонимі болуы мүмкін. Мысалы: *натрий хлориді= ас тұзы*;

Дескриптор жалқы есімі бар тұрақты сөз тіркесі болып табылады. Мысалы: *Цифра заңы* (Мәтіннің барлық сөздерін қолданылу жиілігіне байланысты кему ретімен орналыстырса, онда *n*-ші сөздің жиілігі мұндай тізімде оның реттік нөміріне *n* кері пропорционал болып қалады).

4) Дескрипторда жеке сөздердің мағынасы өте кең болады. Мысалы: *машина* сөзі сөз тіркестерінде: *тігін машинасы, мемлекеттік машина;*

5) Дескриптордың сөз тіркесі компоненттерінің алғашқы әріптерінен құралған көпшілік мақұлдаған аббревиатурасы болады: *Біріккен ұлттар ұйымы – БҰҰ;*

6) Дескрипторды жеке компоненттерге бөлу іздеу үшін маңызды болатын семантикалық байланыстардың жоғалуына әкеліп соқтырады. Мысалы: операциялық жүйе сөз тіркесін бөлу «Unix», «Windows», «Android», «iOS» сияқты тілдік өрнектермен байланыс орнатуға мүмкіндік бермейді

7) Дескриптор кәсіби ұғымдардан құралуы мүмкін. Мысалы: *ықтималды логика, компьютерлік лингвистика;*

8) Дескриптор жалған корреляциялауды болғызбайды. Мысалы: *қоғам туралы ғылым – қоғамтану* терминін бөлу *қоғамдық ғылым* бойынша құжаттардың табылуына әкеп соқтырады;

9) Екі терминнен құралған дескрипторда комбинациялау нәтижесінде біреуінің мағынасы өзгеруі мүмкін. Мысалы: *ұшу алаңы* терминін *ұш алаң* терминдер комбинациясы ретінде қарастыруға болмайды, себебі *ұш* терминіне тән емес пайдалануға әкеледі;

10) Дескриптор қандайда бір толықтықты немесе спецификалық объектіні білдіреді. Мысалы: *сан есім, сөз таптары;*

11) Дескриптордың сөз тіркесі деңгейінде бір немесе бірнеше синонимдері бар немесе сөз тіркесін құрайтын жеке сөздер деңгейінде емес тек сөз тіркесі деңгейінде синонимдері ғана болады. Мысалы: *жартылай өткізгішті триодтар – транзисторлар;*

12) Дескриптор тек жекеше немесе көпше түрде ғана қолданылады. Мысалы: *қазақ тілі, түркі тілдері;*

13) Сөз тіркесінің кейбір элементтерінің осы сөз тіркесінен басқа жерде пайдалануының ықтималдылығы аз. Мысалы: *есін жинады;*

14) Дескрипторда сөз тіркесінің бір элементі басқа элементтің омонимдігін жою мүмкін. Мысалы: *автоматтар: телефон автоматтары, сату автоматтары;*

15) Дескриптор ақпараттық шуды азайтудың бірден бір әдісі болып табылады, яғни, сұранысқа сәйкес келмейтін құжаттарды беруге болады. Мысалы: *«тізбектелген кодты параллелді кодқа түрлендіргіштер» және «параллелді кодты тізбектелген кодқа түрлендіргіштер»* сөз тіркестерінің терминдерін құраушы элементтерге бөліп қарастырсақ, онда *рөлдер* көрсеткіштерінің және осы термин элементтері арасындағы байланыстарсыз оларды ажырату мүмкін емес.

Айтылған шарттарды қанағаттандырмайтын сөз тіркестері компоненттерге бөлінеді.

3. Реляторларды қолдану

Бірмәнді және айқын дескриптор дәл табылмайтын жағдайлар да кездеседі, бұл кезде дескриптор ретінде алынған термин релятормен немесе түсініктемемен бірге беріледі. Релятор ақпараттық іздеу тілінің тәуелсіз лексикалық бірлігі болып табылмайды және ол екі орны да бастауыш позициясында бола алатын екі орындық қатынас болып табылады. Мысалдар: *Есімдік пен үстеу – сөз табы.*

Реляторлар етістіктер мен есімдер кластарындағы сөздердің үлкен тобын құрайды.

Етістіктер арасында келесі қатынастардың көрсеткіштері беріледі:

1) Меңгеру, иелену, бар болу қатынастары: *ие болу, меңгеру, иелену, болу, бар болу.* Мысалы: *кез-келген тілдің әліпбиі болуы керек; Лингвист грамматикалық ережелерді меңгеруі қажет; Дауысты дыбыстар әр түрлі дыбысталуға ие болуы керек; Сөйлемнің соңында нүкте болуы керек; Әрбір сөз түбірге ие.*

2) Бөлшек пен бүтіннің байланысы: *кіру, құрау* (ненің қандай бөлшегі), *сәйкес келу* (ненің неге), *қосу* (өзіне, құрамына) және басқалар, сондай-ақ *бірі* элективті компоненті. Мысалы: *(a)* дыбысы дауысты дыбыстылар құрамына *кіру* керек; *Лексемалар* сөздердің бір бөлшегін *құрайды*; *[a]* дыбысын білдіретін әріпке мәтіннің 20% *сәйкес келеді*; Аффикстер өз құрамына *жұрнақтар* мен жалғауларды *қосу* керек; Қазақ тілі – түрік тілдерінің *бірі*; Түрік тілдерінің қасиеттерінің *бірі* агглютинативтілік болып табылады.

3) Кеңістіктік қатынастар, оның ішінде:

– *Статикалық*: *болу, орнығу, сыйғызу, орналасу, көршілес болу, қысылу, қалу, мекендеу, жетектеу.* Бұл реляторларда локативті модуль міндетті. Мысалдар: Түрік тілдерінде түбір сөздің басында *болады.* Жұрнақ түбірден кейін орнығады; Жұрнақ түбір мен жалғаудың арасында *орналасады.*

Мәтінде реляторлар қызметін өздері релятор болып табылмайтын *тұру*, *жату*, *ілініп тұру* типтес етістіктер атқаруы мүмкін. Мысалы: Нүктелер сөздің соңында *тұруы* мүмкін; Морфология негізінде морфема *жатыр*.

– *Метрикалық: созылу, көсілу*. Мысалы: [r] дыбысының дыбысталуы [t] дыбысынан көбірек *созылады*; Мәтін беттің соңына дейін көсіледі.

4) *Уақыттық қатынастар*, оның ішінде:

– *Реттік: бірдей* (немен), алдында болу (ненің), *еру* (немен), *түсу* (неге), *жатады* (неге). Мысалы: Омонимдердің жазылуы *бірдей*; Сөзде жұрнақ жалғаудың алдында *болу керек*; Түбірден кейін аффикстер *еру керек*; Морфологияда ережелердің көп бөлігі етістіктерге *түседі*. Сөздердің септелмейтін жағдайларына үстеу *жатады*.

– *Метрикалық: созылу, жалғасу, ұзағынан болу және т.б.* Мысалы: [t] дыбысының дыбысталуы 1 секундтан артық *созылмайды*; [m] дыбысының дыбысталуы бірнеше секундқа жалғасуы мүмкін.

5) Алдын-ала келісу қатынастары, оның ішінде:

– Туындау (жағдайдың): *туындау, әкеп соғу* (неге), *мәжбүр ету*, және т.б. Мысалдар: Бірмәнділіктің болмауы релятордың дұрыс анықталмағандығынан *туындаған*; Терминдердің қате классификациялануы шешілмейтін мәселеге *әкеп соғуы* мүмкін; Іздеудің дұрыс болуының шарты тезауруста семантикалық қатынастардың толықтығы мен қарама-қайшылықтың болмауын қамтамасыз етуге *мәжбүрлейді*.

– *Тәуелділік: тәуелді болу, ықпал ету, шарттау, алдын ала анықтау, анықтау*. Мысалы: Тезаурустың дұрыстығы ондағы семантикалық қатынастардың толықтығына *тәуелді*. Тезаурустың құрылымы оның толықтығын қамтамасыз етуге *ықпал етеді*; Қатынастардың аксиомасы олардың жүзеге асырылуын *шарттандырады*; Терминдердің дұрыс классификациялануы тезаурустың дұрыстығын *алдын ала анықтайды*; Сұраныс формасын тезаурустың құрылымы *анықтайды*.

– *Логикалық шығыс: куәлік ету, білдіру, көрсету және т.б.* Мысалдар: Тезаурусты сынау оның дұрыстығына *куәлік етеді*; Тезаурустың бірмәнділігі бір сұраныс бойынша бірден артық жауап алудың мүмкін еместігін *білдіреді*; Тезауруста синонимдердің, омонимдердің, гипонимдердің және гиперонимдердің болуы оның толықтығын *көрсетеді*.

б) *Тұлғааралық (жақын, туыстық) қатынастар: тура келу, болып келу (кімге кім)*: Мысалы: Қазақ тілі басқа түрік тілдеріне қарағанда қарақалпақ тіліне жақынырақ *болып келеді*.

Релятор мәні функционалды емес құралдармен берілуі мүмкін. Тезаурустың жасалуы түрік тілдерінің дамуына *әкелді* → Тезаурустың жасалуы – түрік тілдерінің дамуының себебі → Тезаурустың жасалуынан – түрік тілдері дамыды → Тезаурус жасалғандықтан түрік тілдері дами бастады → Тезаурустың жасалуы – бұл түрік тілдерінің дамуы.

4. Дескрипторлар арасындағы семантикалық қатынастар

Дескрипторлар арасындағы қатынастарды шартты түрде 3 түрге бөлуге болады: «*жақындық қатынастары*», «*иерархиялық қатынастар*» және «*ассоциативті қатынастар*». Бұл қатынастар кейбір формалды қасиеттерге ие болуы мүмкін (*рефлексивті, антирефлексивті, симметриялылық, асимметриялылық, антисимметриялылық және транзитивтік*), олардың мағыналары кейін тезаурустың анықтамасын және терминдер арасындағы қатынастарды жасаған кезде қарастырылады.

Жақындық қатынастарына келесі қатынастар типі жатады:

1) *Антонимия* – дыбысталуы және жазылуы бойынша әртүрлі, бір біріне тура қарама-қарсы мағынаны білдіретін бір сөз табы дескрипторларының семантикалық қатынастарының (антонимдердің) типі. Мысалы, ақиқат – өтірік, мейірімді – зұлым;

2) *Омонимия* – дыбысталуы және жазылуы бойынша бірдей, бірақ мағынасы әртүрлі дескрипторлардың семантикалық қатынастарының (омонимдердің) типі. Мысалы, *эфир* – органикалық зат мағынасында, *эфир* – радио немесе теледидар мағынасында.

3) *Паронимия* – дыбысталуы және жазылуы ұқсас, бірақ мағынасы (мәні) әртүрлі дескрипторлардың семантикалық қатынастарының (антонимдердің, паронимдердің) типі. Мысалы, *адресат* – *адресант*.

4) *Синонимия* – дыбысталуы және жазылуы бойынша әртүрлі, бірақ мағыналары бірдей немесе өте жақын бір сөз табы дескрипторларының семантикалық қатынастарының (антонимдердің) типі. Мысалы, *батыл* – *ержүрек*.

5) *Лексикалық полисемия* – тілдің лексикалық бірлігінің екі немесе одан көп тарихи шартталған немесе мағынасы мен шығуы бойынша өзара байланысқан мағынасының болуы. Мысалы, бидай өрісі, энергетикалық өріс, алгебралық өріс, қызмет өрісі, шайқас өрісі. Өріс сөзінің мағынасын контекст бойынша оңай анықтап алуға болады.

б) *Грамматикалық полисемия* – сөзді (әдетте етістік) бірнеше мағынада қолдануға болады. Мысалы, «Есікке қоңырау шалды» сөйлемінде етістік анықталмаған-жеке мағынада қолданылған, себебі, бұл әрекетті нақты кім жасағаны көрсетілмеген. «Біз туған күн иесіне қоңырау шалдық» сөйлемінде етістік жеке дара мағынада қолданылған, себебі кім қоңырау шалғаны белгілі.

Эквиваленттілік қатынасы *рефлексивті, транзитивті және антисимметрильық* қасиеттерге ие.

Иерархиялық қатынастарға келесі қатынастардың типтері жатады: «*түр-көрініс*» («*жоғары-төмен*») және «*бөлшек-бүтін*».

«*түр-көрініс*» типіндегі қатынас келесі 2 типті қатынасқа ие:

1) *Гипероним* – кең мағыналы дескрипторлардың семантикалық қатынастарының типі, олар объектілердің кейбір ішкі мағынасына қатысты осы объектілердің (заттардың, қасиеттердің, сипаттарының) жалпы түр мағынасын (класс, жиын атауы) білдіреді, яғни, гипероним жалпылаудың логикалық операциясының нәтижесі (жалпыламалық кванторы) болып табылады және заттар (қасиеттер, сипаттар) класының атауы ретінде беріледі және класс элементіне қарағанда әлдеқайда кең мағынаға ие болады. Гиперонимия рефлексивті емес, симметриялы емес және транзитивті қатынас болып табылады. Мысалы, «жануар» термині «ит» терминіне қатысты гипероним болып табылады, ал «ит» термині өз кезегінде «овчарка» терминіне қатысты гипероним болып табылады.

2) *Гипонимия* – басқа әлдеқайда жалпы ұғымдарға қатысты жеке мәнді білдіретін дескрипторлардың (гипонимдардың) семантикалық қатынастарының типі, яғни гипоним шектеу логикалық операциясының нәтижесі болып табылады, класс элементі ретінде беріледі және классқа қарағанда тар мағынаға ие, сонымен қатар гиперонимнің барлық қасиетін еншілеген. Гипонимия рефлексивті емес, симметриялы емес және транзитивті қатынас болып табылады. Мысалы, «овчарка» мәні «ит» ұғымына қатысты гипоним, ал «ит» – «жануар» ұғымына қатысты гипоним болып табылады.

«*бөлшек-бүтін*» қатынасы келесі 2 типті қатынасқа ие:

1) *Меронимия (партонимия)* – бір ұғым басқа ұғымдардың құраушы бөлігі болып табылатын зат есімдер үшін ғана анықталған дескрипторлардың (меронимдердің) семантикалық қатынасының типі, яғни, бұл қатынастың ішінде «элементі болу» және «жасалынған болған» қатынастары көрінеді. Мысалы, автомобиль ұғымының меронимдері қорап, капот, қозғағыш, дөңгелек ұғымдары болып табылады. Меронимияны таксономиядан – «ұқсастыққа» негізделген классификациялаудан айыра білу қажет.

2) *Холоним* – бір ұғымдар басқа ұғымға (ұғымдарға) қарағанда бүтін болып табылатындығын (яғни, басқа ұғымдар біріншісінің құраушы бөлшегі ретінде көрсетілетін) білдіретін зат есімдер үшін ғана анықталған дескрипторлардың (холонимдардың) семантикалық қатынасының типі. Мысалы, автомобиль ұғымы қорап, капот, қозғағыш, дөңгелек ұғымдарына қатысты холоним болып табылады.

Меронимия мен холонимия семантикалық қатынастар сияқты бір біріне өзара кері болып келеді. Тезаурустағы немесе онтологиядағы жазбада меронимдер (партонимдер) әдетте «Part-Of» немесе «АРО» (A-Part-Of), ал холонимдер – «HAS-A» қатынастарымен белгіленеді.

Ұсыныс:

– «*бөлшек-бүтін*» қатынасы тек бір ұғым басқа ұғымға контекстке тәуелсіз түрде кірістірілген жағдайда ғана орнатылуы керек, сол кезде дескрипторлар иерархиясы жасалынуы мүмкін. Мысалы, «*Морфология*», «*Синтаксис*» және «*Семантика*» ұғымдары контекстке тәуелсіз «*Грамматика*» ұғымына кірістірілген болса, ал ол өз кезегінде «*Лингвистика*» ұғымына кірістірілген болса, онда былай жазуға болады:

(*Лингвистика (Грамматика (Морфология) (Синтаксис) (Семантика)))*).

– «*бөлшек-бүтін*» қатынасының 3 түрлі түсіндірмесі болуы мүмкін:

1) *A – B*-ның бөлшегі болып табылады деген *A* кейде *B*-ның бөлшегі бола алады, егер қандай-да бір кезде *A*-ның әрбір мысалы *B*-ның бөлшегі болғанда дегенді білдіреді;

2) *A – B*-ның бөлшегі болып табылады деген *A* кейде *B*-ның бөлшегі бола алады, егер уақытқа тәуелсіз *A* класы *B* класының бөлшегі болады, егер *B*-ның ішкі класы болып табылатын *C*-класы болса және *A*-ның барлық мысалдары *C*-ның бөлшегі болады, сондай-ақ *C*-ның барлық мысалдарында *A*-ның мысалдары бөлшек ретінде болады дегенді білдіреді.

3) *A – B*-ның бөлшегі болып табылады деген *A* сөздігі *B* сөздігіне қосылған дегенді білдіреді, мысалы, сөз тіркестері синтаксисінің онтологиясы сөйлемдер синтаксисінің онтологиясына қосылған.

Лексикалық бірліктер арасында иерархиялық қатынастарды орнатар кезде келесі ұсынымдар беріледі:

– Әрбір жоғарыда тұрған дескриптор төменде тұрған дескриптор жататын базистік семантикалық типке қарасты болуы керек, мысалы, екі дескриптор да (төмендегі және жоғарыдағы)

затты, әрекетті, қасиетті және т.с.с. білдіруі мүмкін. Мысалы, *информатика (пән)* және *компьютер (құрылғы)* дескрипторлары әртүрлі типті ұғымдарға жатады, сондықтан олар иерархиялық қатынастар арқылы байланысуы мүмкін емес. *Компьютер* және *процессор* дескрипторлары «құрылғы» деген бір типті ұғымға жататындықтан олар иерархиялық түрде байланысуы мүмкін.

– Иерархияның көп деңгейін болдырмау үшін берілген пәндік аймақ үшін өте нақты дескрипторларды алып тастау қажет.

– Иерархиялық қатынастар тек контексттен тәуелсіз шынайы болған жағдайда орнатылу керек.

– Түркөріністік қатынастарды сипаттаған кезде «барлығы-кейбірі» тестін қолдану керек. Мысалы, барлық адамдарда ақыл бар, бірақ кейбір адамдар табысты.

Егер дескрипторлар иерархиялық қасиетке ие болса, олардың арасындағы байланыстарды иерархиялық түрдегі қатынас ретінде көрсеткен жөн.

Ассоциативті қатынастар

Дескрипторлар арасында ассоциативті қатынастарды орнатудағы негізгі мақсат индекстеу немесе іздеу кезінде қажет болатын қосымша дескрипторларды көрсету болып табылады. Екі дескриптор арасында ассоциативті қатынас орнатуға болады, егер бір терминді қолданған кезде келесі термин тұспалданатындай болса. Бір термин басқа терминді анықтауда қажетті элемент болуы мүмкін, мысалы, морфема термині сөз терминінің анықтамасының қажетті бөлігін құрайды.

Ассоциативті семантикалық қатынастарға келесі қатынастардың типтері жатады: *себеп-салдар*, *шикізат-өнім*, *үрдіс-объект*, *қасиет-қасиетті иеленуші* және т.б.

Себеп-салдар – бір дескриптордың әрекеті – *себеп* басқа дескрипторды – *салдарды* туындататындай факты орнататын дескрипторлардың (оқиғалардың, үрдістердің) семантикалық қатынасының типі. Бұл ретте бір себеп әртүрлі жағдайда бірдей емес салдар туындатуы мүмкін, бірдей себеп бірдей жағдайда бірдей салдар туындатуы мүмкін.

Себеп-салдар қатынасы басты және басты емес; жай және құрама; жалпы және ерекше; жеке және көпше; ішкі және сыртқы; объективті және субъективті; ақпараттық және энергетикалық; динамикалық және статикалық; бірфакторлы және көп факторлы; жүйелі және жүйелі емес.

Шикізат-өнім – берілген дескриптор – шикізаттан алдағы уақытта өңделіп, басқа дескриптор – өнімді тудыратын факты орнататын дескрипторлардың семантикалық қатынасының типі, бұл ретте кез келген өнім шикізат болып, басқа өнімге айнала алады. Мысалы, әріптерден сөз құралады, ал сөздерден сөйлем құрастыруға болады, ал сөйлемдерден мәтін жасауға болады.

Үрдіс-объект – дескриптор – үрдіс қандай да бір нәтижеге жету үшін шынайы немесе ойдан шығарылған әлемде бір бүтінді білдіретін басқа дескриптордың – объектінің келесі қалпын ауыстыруды көрсететін факты орнататын дескрипторлардың семантикалық қатынасының типі.

Мысалы, берілген сөзге жұрнақты жалғау үрдісі жаңа сөзді жасауға мүмкіндік береді, ал жалғауды қосу үрдісі – оның формасын өзгертеді.

Қасиет-қасиетті иеленуші – берілген дескриптордан – қасиеттен, яғни басқа дескриптордың атрибутын көрсететін – қасиетті иеленуші, осы қасиетке ие объектілерді көрсететін факты орнататын дескрипторлардың семантикалық қатынасының типі.

Терминдер арасындағы семантикалық қатынастар мағынасы тезаурустың анықтамасын және терминдер арасындағы қатынастарды жасау кезінде берілетін келесі қасиеттерге ие:

– Антонимия және омонимия жақындық қатынастары рефлексивті емес, симметриялы емес және транзитивті емес қасиеттерге ие, ал паронимия және синонимия – рефлексивті, симметриялы және транзитивті қасиеттерге ие.

– Иерархиялық семантикалық қатынастар транзитивті және антисимметриялылық қасиеттерге ие, ол ақпараттық іздеудің тиімділігін жоғарылату мақсатында индекстеу артық болғанда пайдаланылады;

– Ассоциативтіліктің қатынастары рефлексивті, транзитивті және антисимметриялылық қасиеттерге ие.

Осы қарастырылған мәселелерді ескере отырып жасалған тезаурус мысалынан үзінді:

Дескриптор: Есімді тіркес

UniTurk: NP

Анықтамасы: Басыңқы бөлімі есім сөздер болатын сөз тіркесі

Гипероним: Еркін сөз тіркесі

Гипоним: Есімді меңгеру, Есімді матасу, Есімді қабысу

Мероним: Зат есім, Сын есім, Есімдік, Сан есім

Қорытынды

Бұл жұмыста тезаурусқа, сонымен қатар оның түсіндірме сөздіктен айырмашылығына ерекше көңіл бөлінеді. Ол адамға бағытталған сияқты индекстеу мен ақпараттық іздеу жүйелеріне бағытталуы мүмкін. Ақпараттық-іздеу тезаурусы мәтінді табиғи тілден дескрипторлы ақпараттық-іздеу тіліне лексикалық бірліктерді сәйкес дескрипторлармен бірмәнді ауыстыру жолымен аударуда және дескрипторлар арасындағы парадигматикалық байланыстарды көрсетуде пайдаланылады. Тезаурус мәтіндерді автоматты өңдеу сияқты қосымшаларда қолдануға арналған.

Осы мақалада сипатталған тезаурусқа қойылатын талаптарды қолдана отырып түрік тілдері (қазақ, қырғыз, татар, түрік және өзбек) грамматикасының электронды тезаурустары жасалады.

Мақала 2018 жылғы 12 наурыздағы №132 келісім шарт бойынша ЖТН АР05132249 «Көптілді іздеу және білімдерді шығару жүйелерін құруға арналған түркі тілдерінің электрондық тезаурустарын әзірлеу» жобасы аясында дайындалды.

Пайдаланылған әдебиеттер тізімі:

- 1 Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011. – 512 с.
- 2 ISO 2788-1986. Documentation – Guidelines for the establishment and development of monolingual thesauri. Ed. 2.
- 3 ISO 5964-1985. Documentation - Guidelines for the establishment and development of multilingual thesauri, IDT (Revised by: ISO/DIS 25964-1 Under development).
- 4 ГОСТ 7.25-2001. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления (введен в действие с 1 июля 2002 г.)
- 5 ГОСТ 7.24-2007. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению
- 6 ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (Periodic Review).
- 7 Балақаев М., Сайрамбаев Т. Қазіргі қазақ тілі. Оқулық. Өңд. және толықт., 3-бас. – Алматы: Санат, 1997. – 239 б.

References

- 1 Lukashovich N.V. (2011) Tezaurusy v zadachah informacionnogo poiska [Thesauruses in the problems of information search]. Izd-vo Moskovskogo universiteta. 512. (In Russian)
- 2 ISO 2788-1986. Documentation – Guidelines for the establishment and development of monolingual thesauri. Ed. (In English)
- 3 ISO 5964-1985. Documentation - Guidelines for the establishment and development of multilingual thesauri, IDT (Revised by: ISO/DIS 25964-1 Under development). (In English)
- 4 (2002) GOST 7.25-2001. Sistema standartov po informacii, bibliotechnomu i izdatel'skomu delu [The system of standards for information, library and publishing]. Tezaurus informacionno-poiskovyj odnojazychnyj. Pravila razrabotki, struktura, sostav i forma predstavlenija (vveden v dejstvie s 1 ijulja 2002 g.). (In Russian)
- 5 (2007) GOST 7.24-2007. Sistema standartov po informacii, bibliotechnomu i izdatel'skomu delu [The system of standards for information, library and publishing]. Tezaurus informacionno-poiskovyj mnogojazychnyj. Sostav, struktura i osnovnye trebovanija k postroeniju. (In Russian)
- 6 ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (Periodic Review). (In English)
- 7 Balakaev M., Sajrambaev T. 1997 Kazirgi kazak tili. Okulyk. Ond. zhane tolykt. 3 bas. Almaty: Sanat. 239. (In Kazakh)