

**СЎЗ САНЪАТИ
ХАЛҚАРО ЖУРНАЛИ**

2 - МАХСУС СОН

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИСКУССТВО СЛОВА**

СПЕЦИАЛЬНЫЙ ВЫПУСК - 2

**INTERNATIONAL JOURNAL
OF WORD ART**


SPECIAL ISSUE - 2



СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ
МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА
INTERNATIONAL JOURNAL OF WORD ART

Манзура Абжалова Абдурашетьовна
Навоий давлат кончилиқ институти,
Филология фанлари доктори
Искандаров Отабек Исмоилович
Навоий давлат кончилиқ институти,
Илмий изланувчи

**МИЛЛИЙ КОРПУСИ МАВЖУД БЎЛМАГАН ТИЛЛАРНИНГ
ЛИНГВИСТИК ДАСТУРЛАРИДА ОМОНИМЛАРНИ ТАҲЛИЛ
ҚИЛИШ ТЕХНОЛОГИЯСИ**

 <http://dx.doi.org/10.26739/2181-9297-2020-SI-2-13>

АННОТАЦИЯ

Омоним шакллар ҳамиша тилшунослик марказидаги ҳодисалардан бири сифатида қаралган. Шу боис матнда учрайдиган омошакллар лингвистик корпусга боғлиқ ҳолда рус ва хориж тилшунослигида алоҳида тадқиқ қилинади. Мазкур мақолада дунё тилшунослигидаги лингвистик таъминотлар асосида яратилган лингвистик дастур ва тизимларда омонимшаклларни таҳлил қилиш усул ва моделларидан Брилл усули, яширин Марков модели, моделлар модификацияси таҳлил қилинди ва уларнинг тил корпусига таянишига, миллий корпуснинг тил материалларини қайта ишлашдаги аҳамиятига муносабат билдирилди. Мақолада ўзбек тили Миллий корпуси яратилмагани боис ўзбек тилидаги матнларни таҳрир ва таҳлил қилиш дастури учун N-gramмга боғлиқ равишда яратилган омонимларни таҳлил технологияси берилди.

Таянч сўзлар: омоним, корпус, модель, тег, ID.

Manzura Abjalova Abdurashetovna
Navoi State Mining Institute,
Doctor of Philosophy in Philology
Iskandarov Otabek Ismailovich
Navoi State Mining Institute,
Scientific researcher

**TECHNOLOGY OF ANALYSIS OF HOMONYMS IN
LINGUISTICS PROGRAMS OF LANGUAGE WHICH DOES NOT**

NATIONAL KORPUS

ANNOTATION

Homonymous forms have always been considered one of the language events in the center of linguistics. In this regard, omofoms found in the texts are separately studied in Russian and European linguistics in conjunction with the linguistic corpus. This article analyzes methods and models in linguistic programs and systems based on linguistic software in world linguistics, such as the Brill method, hidden Markov model, modification of models and their connection with the language corpus, and gives an attitude to the importance of the national corpus in language processing. The article presents the technology for the analysis of homonyms of the dependent N-gram, which was developed for a program for editing and analyzing texts in the Uzbek language that are not based on the National Corpus.

Keywords: homonym, corpus, model, tag, ID.

Манзура Абжалова Абдурашетовна

Навоийский государственный горный институт,
Доктор философских наук в области филологии

Искандаров Отабек Исмаилович

Навоийский государственный горный институт,
Научный сотрудник

ТЕХНОЛОГИЯ АНАЛИЗА ОМОНИМОВ В ЛИНГВИСТИЧЕСКИХ ПРОГРАММАХ ЯЗЫКА , КОТОРЫЕ НЕ ИМЕЮЩИХ НАЦИОНАЛЬНОГО КОРПУСА

АННОТАЦИЯ

Омонимические формы всегда считались одним из языковых событий в центре лингвистики. В связи с этим, омоформы, встречающиеся в текстах, отдельно исследуются в русской и европейской лингвистике во взаимосвязи с лингвистическим корпусом. В данной статье анализируются методы и модели в лингвистических программах и системах, основанных на лингвистическом программном обеспечении в мировой лингвистике, как метод Бриля, скрытая марковская модель, модификация моделей и их связка с языковым корпусом и дано отношение к важности национального корпуса в языковой обработке. В статье представлена технология анализа омонимов зависимой N-граммы, которая разработана для программы редактирования и анализа текстов на узбекском языке, не опирающихся на Национальный корпус.

Ключивые слова: омоним, корпус, модель, тег, ID.

Маълумки, омонимлар ёзилиши бир хил фонемалардан иборат шакллар ҳисобланади. Шу боис омонимияда ифода плани етакчи Раҳматуллаев, 1984:5 . Омоним ҳодисада ясама лексема ва туб лексема, шунингдек, уларга грамматик маъно ифодаловчи аффикс қўшилганда воқе бўладиган шаклий тенг келиш ҳам қамраб олинади. Масалан, терим сўзшакли тер - феъл лексемасига - им лексема ясовчисининг бирикиши натижасида ҳосил қилинган лексеманинг бирлик, бош келишиқ формасига тенг.

Аниқроқ бўлиш учун, бир неча турдаги омонимларни ажратиш кўрсатиш мумкин:

- 1) нутқнинг бир қисми ёки луғатдаги шакли (лемма) фарқланади;
- 2) баъзи морфологик хусусиятлар, масалан, келишик, эгалик ёки сонда (бой морфологияли тилларда мавжуд) фарқланади;
- 3) бир-биридан фақат мазмун жиҳатдан фарқланади (бу ҳолат семантик фарқланишда учрайди).

Жаҳон тилшунослигида XX асрнинг 50-йилларидан эътиборан матнни электрон ҳисоблаш машиналари воситасида таҳрир қилиш муаммоси ўрганила бошланди, натижада замонавий ахборот технологиялари дастурлари ҳамда табиий тилнинг грамматик қонуниятларига асосланган матн таҳрир ва таҳлилларининг лингвистик таъминоти юзага кела бошлаган. Дунё тилшунослигида бундай лингвистик таъминотлар асосида яратилган лингвистик дастур ва тизимларда омонимшаклларни таҳлил қилиш усул ва моделлари (Брилл усули, Яширин Марков модели, моделлар модификацияси) ишлаб чиқилган бўлиб, улар фақатгина миллий тил корпуси орқали ишлайди.

Тилнинг морфологик сатҳини ўзида қамраган универсал тизим (корпус) мавжуд эмаслиги сабабли кўпгина тилларда турланган сўзшакллари ёки сўз туркумлари доирасидаги сўзларнинг аффикслар билан бирикишида лингвистик таъминотга киритилган морфологик луғатларга мурожаат қилинади. Шу боис турли хил луғатлар учун омонимлик (ноаниқлик) тушунчаси биров фарқ қилиши мумкин.

Морфологик луғатлар билан бир қаторда корпус, деб аталадиган тўлиқ матнлар тўплами мавжуд Захаров, www.dialog-21.ru/media/2138/zakharov.pdf (ўзбек миллий корпуси яратилмаган). Корпус таъминотининг шакллантирилишига қараб омоним шаклларнинг нафақат морфологик, балки семантик ва синтактик хусусиятлари ҳам намоён бўлади <http://opencorpora.org>. Ахборот олишнинг қулай имкониятлари яратилиши натижасида корпус тил ҳақидаги мукамал маълумот манбаига айланади, масалан, омоним шаклларнинг энг кенг тарқалган тури статистикасини билиш мумкин бўлади.

Бугунги кунда лингвистик тадқиқот ва амалий топшириқлар ечими учун тил корпуслари замонавий тилшуносликнинг инкор этиб бўлмас иш қуролига айланди, чунки корпус турли лингвистик топшириқларни ечишга хизмат қилади Недошивина, 2006:26; Хамроева, 2018; Кобрицов, 2004 .

Корпус муайян электрон манбага лексик, морфологик, грамматик, семантик белгилар асосида ишлов берилганлиги сабабли лингвистик тадқиқотларда фойдаланиш учун жуда қулай имкониятни беради. Ш.Хамроева таъкидлаганидек, тил корпуси электрон кутубхонадан фарқли равишда муайян тилни тадқиқ қилиш, ўрганиш ва ўргатиш учун зарурий, фойдали ва қизиқарли матнлар тўпланишини назарда тутди Хамроева, 2018 .

Турли корпуслардаги омонимлик турлича бўлади. Бу морфологиянинг нафақат ҳар хил моделларидан фойдаланилишига, балки матнлар услуби ва маълумотлар миқдорига ҳам нисбатан боғлиқдир. Рус тили корпусига бирлаштирилган матнлар таркибида омонимликка эга сўзшакллар кўп учрайди. Шу боис хорижий ва рус тилшунослигида омонимликни бартараф этиш масаласи ("снятия омонимии") алоҳида тадқиқ қилинади 4,5,7,8,9,14 . Қуйида омоним сўзларшаклларни матн таркибида аниқлашнинг кенг қўлланиладиган усуллари ҳақида сўз юритилди.

Грамматикага оид манбалардан маълумки, омонимияни йўқ қилиш (яъни ноаниқликни бартараф этиш)нинг деярли барча усуллари икки гуруҳга бўлинади Рысаков, <http://samag.ru/archive/article/3059> :

1. Грамматик меъёрларга асосланган усуллар. Ўз навбатида, улар қуйидагича

гуруҳланади:

- технологияларнинг аралашувисиз қўлда киритилган методлар;
- қоидаларни автоматик яратиш усуллари.

2. Статистикага асосланган усуллар.

Ушбу гуруҳларнинг ҳар бирида афзаллик ва камчиликлар мавжуд. Шундай ҳолат бўладики, ҳар иккала гуруҳга тегишли методлар комбинациясидан юзага келган метод яхшироқ натижани кўрсатиши мумкин. Бундай гибрид метод билан танишишдан олдин ҳар бир турнинг вакиллари кўриб чиқиш мақсадга мувофиқдир.

Брилл усули. Қоидаларни автоматик генерация қилиш методининг яққол намунаси - америкалик лингвист Эрик Брилл методидир Brill, 1995:543-565 . Иш услуби қуйидагича:

1. Бошлаш: Ҳар бир сўз ўша сўзнинг тез-тез ишлатиб туриладиган теги (нутқнинг асл қисмида) билан боғланиши лозим. Номалум сўзлар от сўз туркуми сифатида қабул қилинади. Бошлаш босқичидан нафақат ўрганиш жараёни бошланади, балки омонимларни бартараф этиш методи ҳам ҳаракатга келади.

2. Тез-тез учрайдиган хатолик учун ўзгариш (қайта ишлаш) қоидасини яратиш.

3. Исталган минимал хатога эришгунча иккинчи босқични такрорлаш.

Ўтказиш (қайта ишлаш) қоидаларида "эски тег, янги тег, шарт" тўпламлари ўз ифодасини топади ва қоидада белгиланган шарт бажарилганда эски тег янгиси билан алмаштирилади. Ушбу методнинг камчилиги қоидалар сонининг кўпайиши билан аниқлик даражасининг пасайишида кузатилади Brill, 1992:154 , бу Паретонинг принципига мос келади: 80% ҳаракат 20% натижани таъминлайди. Шу билан бирга, принцип аксинча жараёнда ҳам ишлайди: бошлаш босқичининг фақат битта қадамни бажариш натижасида омонимликни бартараф этишда юқори аниқликка эришиш мумкин. СинТагРус корпусида ўтказилган тест натижаларидан маълум бўлдики, ушбу метод ҳар бир сўзнинг гапдаги ўрнини 97,4%, морфологик хусусиятларнинг тўлиқ мажмуини 87,6% аниқлик билан топиш имконини беради.

Шаклдош сўзларни бартараф этиш методининг фойдали ишини баҳолаш учун статистикани барча сўзлар билан эмас, балки фақат омонимлар бўйича ҳисоблаш керак. Бундан аниқ бўладики, нутқнинг бир қисмини белгилаш вазифаси корпуснинг 24,10 %га тўғри келади, гап бўлаги ва морфологик хусусиятларни аниқлаш эса 51,94 %ни ташкил қилади. Брилл усулининг қисқартирилган версияси бўлган юқорида айтилган частота методи учун фойдали аниқлик 89,1% бўлса, гап бўлаги ва морфологик хусусиятларни аниқлаш кўрсаткичи 76,2% ни ташкил қилади.

Яширин Марков модели. Баум Л.Е. ва унинг ҳамкасблари томонидан ишлаб чиқилган Baum, 1968:211-227 мазкур модель омонимликни аниқлашнинг статистик методи статистик жараёнда юзага келадиган барча вариантлар эҳтимоллигини ҳисобга олишга ёрдам беради. Масалан, маълум бир матнда от туркумига оид сўзлар боғловчига нисбатан тез-тез ва кўп учраса унда айна контекстда мавжуд омоним катта эҳтимоллик билан боғловчи эмас, от туркумига оид сўз бўлади, кейинги эҳтимолликда боғловчи сифатида ҳисобга олинади. Контекстни тавсифлаш учун N-граммадан фойдаланилади. N-грамма - матнларга автоматик ишлов беришда кенг қўлланиладиган математик ҳисоб воситасидир. Ўзбек компьютер лингвистикасида С.Ризаев ҳарф бирикмаларини биграмм, триграмм терминлари билан ифодалаган Ризаев, 2006:18 .

N-грамма - сўзлар ёки теглар каби N-идентификатор элементларнинг кетма-кетлигини ифодалайди. Икки элемент кетма-кетлиги - биграмма, уч элемент

кетма-кетлиги эса триграмма, дейилади. Масалан, олд кўшимча+от ҳолати биграммага мисол бўлади.

Омонимликни аниқлашнинг оддий статистик методи ва бошқа шу каби методларнинг тавсифини келтириш учун қуйидаги усуллар ишлатилади:

– w_i – жумладаги i -ўринда жойлашган сўз, t_i – ушбу сўзнинг идентификатори (теги).

– $D(w) = \{t_1^w, t_2^w, \dots, t_k^w\}$ w сўзининг барча мумкин бўлган белгилар мажмуи. Ушбу маълумотларни морфологик луғат ёрдамида олиш мумкин. Агар сўз луғатда бўлмаса уни Брилл усулида бажарилганидек, от сўз туркуми сифатида ҳисоблаш мумкин, аммо лингвистик таъминот ишончли бўлиши учун барча мумкин бўлган тегларни қўйиб чиқиш керак.

– C – корпусдаги муайян ҳолатлар сони (n -грамм). Бунда $C(t)$ – t теглар сони; ва $C(t_1, t_2)$ – биграммалар сони (t_1, t_2) .

– $C_t(w, t)$ – t тегли w сўзлар сони.

– $F(w, t)$ – w сўзида t теги мавжудлиги эҳтимоли. Тавсифлар қуйидаги формула бўйича ҳисобланади:

$$F(w, t) = \frac{C_t(w, t)}{C(t)}$$

– $P(t_i|t_{i-1})$ – бу t_{i-1} тегидан кейин t_i тегининг келиш эҳтимоллик ҳолати. Бунда $i = 1$ бўлганда t_i теги гапда биринчи тег ҳисобланади. Ҳисоблаш формуласи қуйидагича:

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})}$$

Яширин Марков моделига асосланиб, омонимликни аниқлашнинг статистик методи ёрдамида ишлаш натижасида n узунликдаги жумлада $T_i \in D(w_i)$ бўлганда $T = \{T_1, T_2, \dots, T_n\}$ тегларнинг эҳтимолий кетма-кетлиги топилади [http://www.academia.edu/15517740/Анализ_статистических_алгоритмов_снятия_морфологической_омонимии_в_русском_языке].

Моделлар модификацияси. Лингвистик корпусдаги статистикага асосан эҳтимоллик ҳисоб-китобида баъзи кам учрайдиган теглар жуфтлиги учун P эҳтимоллиги нолга тенг бўлади. Бу арзимасдек туюлган хатолик қолган эҳтимолликлардан қатъий назар, кетма-кетлик натижасининг нолга тушишига олиб келиши мумкин. Ушбу камчиликни бартараф этиш учун статистикада силлиқлаш жараёни [Рысаков, <http://samag.ru/archive/article/3059>] бажарилади. Оддий мисол, P функциясида Лапласнинг силлиқланишида P нинг қиймати нолдан каттароқ бўлиши талаб қилинади:

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i) + 1}{C(t_{i-1}) + 1}$$

Албатта, силлиқлаш – қулай усул эмас, шунинг учун унга жуда кўп таянилмайди. Лингвистик корпус ҳажми қанчалик катта бўлса тўпланган маълумотларга нисбатан кўпроқ ишонч ва кам учрайдиган лингвистик ҳолатларни ўтказиб юбориш эҳтимоли камроқ бўлади [Клышинский, 2015: 556-557]. Одатда корпуснинг ҳажми чекланган, аммо статистика йиғиш учун хусусиятларни танлаш эркинлиги мавжуд. Биграммалар ўрнига триграммалар ҳисобга олиниши мумкин. Бу ҳолда кидирув формуласи қуйидаги шаклни олади:

$$P(t_i|t_{i-1}, t_{i-2}) = \frac{C(t_{i-2}, t_{i-1}, t_i)}{C(t_{i-2}, t_{i-1})} \text{ ҳолатида}$$

$$\operatorname{argmax}_T \prod_{i=1}^n P(T_i | T_{i-1}, T_{i-2}),$$

бўлади.

Лингвистик таъминотни яратишда омонимларни аниқлаш методида текшириладиган белгиларнинг тил хусусиятларига тиниш белгиларини ҳам қўшиш мумкин, аммо бу ҳолатда тиниш белгилари N-граммада ҳисобга олинishi муаммолигича туради. Масалан, агар пунктуация алоҳида тег сифатида ҳисобга олинса биграмма статистикасида орасига тиниш белгисининг қўйилиши натижасида жуфт теглар алоҳида-алоҳида теглар, деб қабул қилинади. Теглардаги кетма-кет жуфтликни сақлаб қолиш учун биграмма (тег, тег)ни триграмма (тег, пунктуация, тег)га алмаштириш мақсадга мувофиқдир. Бунда қидирув ҳолати қуйидагича тус олади:

$$\operatorname{argmax}_T \prod_{i=1}^n P(T_i | T_{i-1}, pm_{i-1}),$$

$$P(t_i | t_{i-1}, pm_{i-1}) = \frac{C(t_{i-1}, pm_{i-1}, t_i)}{C(t_{i-1})},$$

pm_i – i сўздан кейинги тиниш белгиси. Бу ерда ва олдинги мисолларда маҳраж бошқа кўринишга эга бўлиши мумкин: i сўзнинг исталган шаклида маҳраж барча занжирлар учун бир хил бўлса максимал изланишга таъсир қилмайди. Частотали моделда бўлгани каби сўзнинг белгиси сифатида фойдаланиш мумкин.

Амалиёт шуни кўрсатдики, яширин Марков моделидаги теглар частотаси тўғрисидаги маълумотлардан аниқликка эришиш мумкин. Тиниш белгиларсиз биграммаларни ишлатганда қидириш формуласи қуйидаги ҳолатда бўлади:

$$\operatorname{argmax}_T \prod_{i=1}^n P(T_i | T_{i-1}) \cdot F(w_i, T_i)$$

барча сўзларни алоҳида гуруҳларга ажратиб, ҳисоб-китобни янада аниқроқ қилиш мумкин. Буни амалга ошириш учун $G(w)$ функцияси киритилади ва ҳар бир сўз ўзи тегишли бўлган гуруҳ идентификатори билан боғланади. Гуруҳларнинг киритилиши билан P функцияси қуйидаги шаклга келади:

$$P_G(t_i | t_{i-1}; g_i) = \frac{C_{g_i}(t_{i-1}, t_i)}{C(t_{i-1})},$$

бу ерда $g_i = G(w_i)$, ва C_g охириги сўз g гуруҳига тўғри келадиган n -грамм сонини билдиради. Сўзларни гуруҳлашда қўлланиладиган методнинг танланишига боғлиқ равишда методнинг аниқлигини ошириш ёки камайтириш ҳам мумкин. Муаллиф томонидан таклиф қилинган модификацияда ҳар бир гуруҳда сўз қабул қилиши мумкин бўлган мос келувчи тег варианты рўйхати, яъни $G(w) = D(w)$ мавжуд. Бунинг натижасида қидириш формуласи тубандагича бўлади [Рысаков, <http://samag.ru/archive/article/3059>]:

$$\operatorname{argmax}_T \prod_{i=1}^n P_G(T_i | T_{i-1}; D(w_i))$$

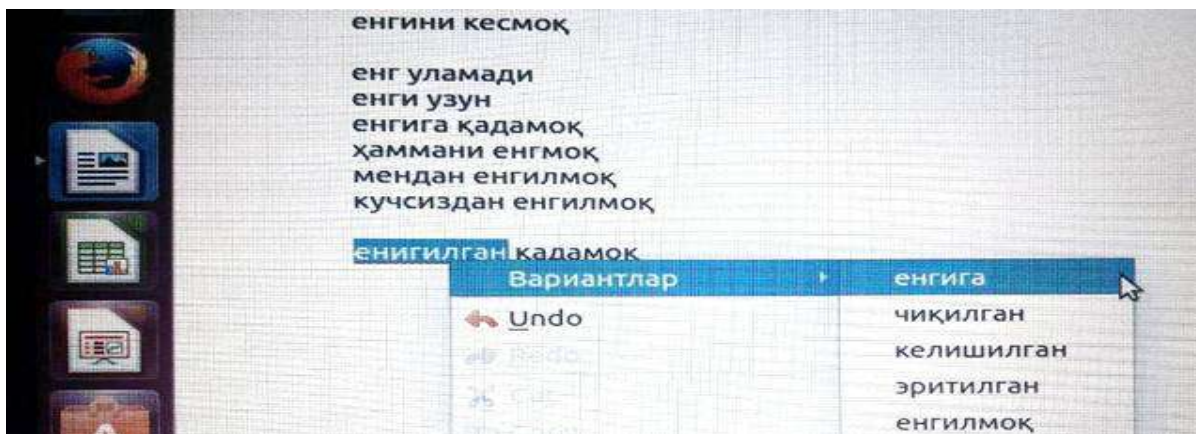
Таъкидлаш жоизки, манбалари кўрсатилган изланишларда омонимлик ҳодисаси корпус лингвистикасига боғлаб тадқиқ қилинган. Ваҳоланки, кўплаб тилларнинг миллий корпуси хануз яратилган эмас. Шунга биноан Брилл усули, яширин Марков

моделли, моделлар модификацияси каби методларидан миллий корпусига эга бўлмаган тиллар матнидаги омонимларни таҳлил қилишда қўллаб бўлмайди.

Ўзининг Миллий корпусига эга бўлмаган ўзбек тилидаги матнларни автоматик таҳлил қилувчи дастурларда омонимшаклларни таҳлил қилиш усулини яратиш эҳтиёжи ошди. Шу боис ўзбек тилидаги матнларни таҳлил ва синтез қилувчи дастур (DГУ 06645)да матнларни таҳлил қилиш жараёнида омоним шаклларни таҳлил қилиш учун N-grammга боғлиқ равишда омоним-таҳлил технологияси яратилди. Мазкур усул миллий тил корпусига эга бўлмаган барча тилларни автоматик синтез ва таҳлил қилувчи дастурларда қўллansa иш самарадорлиги ошади.

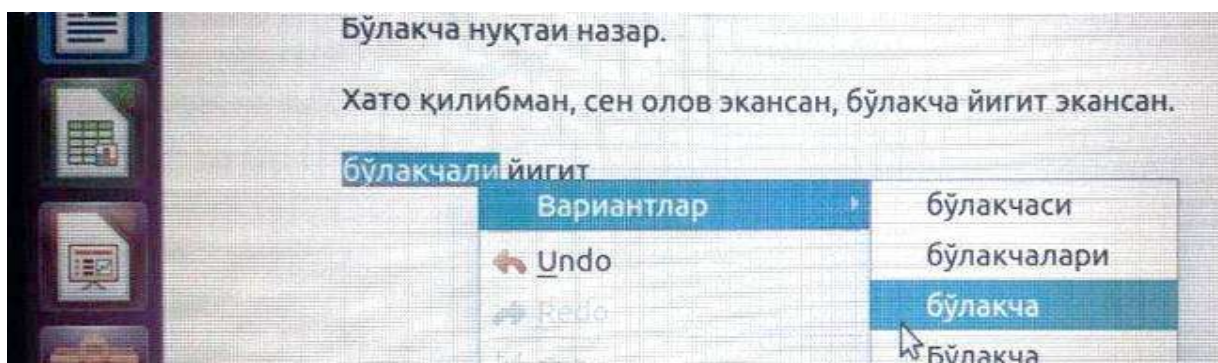
Технологиянинг иш тартиби: Бунинг учун ўзбек тилидаги омонимлар ЛТга ID билан киритилди, бу жиҳат контекстда омонимларнинг грамматик шаклланишини таҳлил қилиш имкониятини беради. Таҳлил аниқлигини ошириш мақсадида омонимшаклнинг туркуми аниқланди ва унга рамзий қиймат берилди (BGU 00381-рақамли гувоҳнома). Шундан сўнг омонимлар базаси (List of homonym) сўз бирикмалари модели (Models of CompWords)га боғланди (List of homonym=> Models of CompWords). Натижада киритилган М – омоним ўзи билан келган N – сўзга боғлиқ равишда ЛТдаги М – омонимнинг ID рақамига тўғри келувчи Q – аффиксларни олади. Яъни, M[id]+Q[id]. Мазкур усулни мисол орқали таҳлил қиламиз: $енг_1 \rightarrow N/от$ (кийимнинг қўлни қоплаб турадиган қисми), $енг_2 \rightarrow V/феъл$ (ғолиб келмоқ, устун келмоқ), $енг_3 \rightarrow V/феъл$ (ейишга ундовчи буйруқ феъли) омонимшакллар матнда синтактик модулда яратилган сўз бирикмалари моделидаги сўз туркумлари рамзларига мос равишда ўзидан кейин келган сўз (биграмма – икки белги кетма-кетлиги) билан синтактик боғланишда қабул қиладиган грамматик воситаси таҳлил қилинади. Демак, бу ерда N ва V билан боғлиқ моделларга мурожаат қилинади: $N_{\text{сом-ни}} + V$ (*енгини кесмоқ*), $N_{\text{сом-га}} + V$ (*енгига қадамоқ*), $\text{Pron}_{\text{ни(дир)}} + V$ (*ҳаммани енгмоқ*), $\text{Pron}_{\text{дан(дир)}} + V$ (*мендан енгилмоқ*), $\text{Adj}_{\text{дан}} + V$ (*кучсиздан енгилмоқ*) ва ҳ.к.

Омоним сўз	Туркуми	туркум рамзи	Рус тилидаги таржимаси	Изоҳи	Мисол
енг	от	N	рукав	кийимнинг қўлни қоплаб турадиган қисми	Енгли, енгсиз, енглик, енгча; Азиз ошни гулдор сопол лаганга сузиб келди. Учовлон чордона қуриб олиб, енгни шимардилар. Ҳамид Ғулом, Бинафша атри.
енг	феъл	V	побеждать, одолевать, побороть	ғолиб келмоқ, устун келмоқ	Сени иш енгмасин, сен ишни енг. Мақол.
енг	феъл	V	ешь	ейишга ундовчи буйруқ феъли	Олинг! Овқатдан енг!



"бўлакча" омонимшаклининг таҳлил қилиниши:

Омоним сўз	Туркуми	туркум рамзи	Рус тилидаги таржима си	Изоҳи	Мисол
бўлакча I	от	N	кусочек	бўлакнинг г парчаси, кичкина бўлак	Новвот бўлакчалари.
бўлакча II	сифат	Adj	иной	кутилган дан фарқли; бошқача, ўзгача	Бўлакча муносабат. Бўлакча нуқтаи назар. Хато қилибман, сен олов экансан, бўлакча йигит экансан. Ойбек, Қутлуғ қон.



Гуруҳлардаги сўзларнинг боғланиши ва ўзаро муносабатлари битишувли, мослашувли ва бошқарувли бирикмалар моделлари жадвалларида келтирилди.

Битишув йўли билан боғланиш модели:

турдош от - N_{com}, феъл - V, сифат - Adj, сон - Num, равиш - Adv, олмош - Pron.

№	Битишувли бирикмалар модели	мисоллар
1.	Ncom+ Ncom	тилла узук, ижодкор талаба
2.	Ндаги+ Ncom	ишдаги шижоат, ходимдаги масъулият
3.	Ндаги+ Num	қарордаги саккиз..., иловадаги учинчи жадвал
4.	Advдаги+ Ncom	қуйидаги мисоллар, юқоридаги фикрлар
5.	Num + Ncom	иккита шартнома, тўрт кафедра
6.	Pron + Ncom	барча иш, қандай тажриба
7.	Adj + Ncom	масъулиятли ходим, обод маҳалла
8.	Adv + Ncom	кўп иш, яккама-якка жанг, тўхтовсиз кураш
9.	Vh + Ncom	тақдирлаш куни, нишонлаш куни
10.	Ндай/дек+ Adj	кундек равшан, пахтадек момик
11.	Vs + Adj	(таълим) олаётган талабалар, фидойи ёшлар, тузалган бемор
12.	Adj + V	тўғри сўзламоқ, яхши бормоқ
13.	Adv + V	эскича қараш, ўз-ўзидан гапирмоқ, янгича ёндашув
14.	Vr + V	билиб-билмай гапирмоқ, яқунлангунча улгуриш

Мослашувли бирикмалар модели яратилди:

№	Мослашувли бирикмалар модели	мисоллар
	Ncom-нинг+ Ncom-эгаллик афф.	қарорнинг банди, буйруқнинг ижроси
	Ncom-нинг+ Adjэгаллик афф	ишнинг каттаси, муқованинг юпқаси
	Ncom-нинг+ Numэгаллик афф	столнинг биттаси, ҳужжатнинг иккитаси
	Adjнинг + Ncom-эгаллик афф.	каттанинг ўрни, кичикнинг ҳурмати
	Adjнинг + Adjэгаллик афф.	каттанинг оласи
	Numнинг + Numэгаллик афф	бешнинг ярми, иккитанинг бири
	Ncom-нинг+ Vhэгаллик афф (Vh – ҳаракат номи)	ректорнинг қайтиши, мавзунинг ёритилиши, ижронинг кечиктирилиши
	Vhнинг + Ncom-эгаллик афф.	изланиш(нинг) вақти, йиқилишнинг оқибати
	Vhнинг + Vhэгаллик афф.	олмоқнинг бермоғи
	Ncom-нинг+ Vsэгаллик афф (Vs– сифатдош)	олимнинг излангани, вақтнинг ўтгани
	Vsнинг+ Ncom-эгаллик афф.	ўқиганнинг тафаккури, сўраганнинг айби

Бошқарувли бирикмаларда моделлар миқдорининг ошиб бориши эътиборга олиниб, бундай бирикмаларни қуйидаги умумий модель орқали ифодалаш

мақсадга мувофиқ:

$\left. \begin{array}{l} \text{Ncom} \\ \text{Adj} \\ \text{Num} \\ \text{Pron} \\ \text{Vh} \end{array} \right\} +\text{-ни} +\text{V}$	$\left. \begin{array}{l} \text{Ncom} \\ \text{Num} \\ \text{Adj} \\ \text{Num} \\ \text{Adj} \\ \text{Vh} \end{array} \right\} +\text{-га} +\text{V} \left\{ \begin{array}{l} \text{N} \\ \text{Pron} \\ \text{Vh} \end{array} \right.$	$\left. \begin{array}{l} \text{Ncom} \\ \text{Num} \\ \text{Adj} \\ \text{Num} \\ \text{Pron} \\ \text{Vh} \end{array} \right\} +\text{-да} + \left\{ \begin{array}{l} \text{N} \\ \text{V} \\ \text{Adj} \\ \text{Vh} \end{array} \right.$
--	---	---

$\left. \begin{array}{l} \text{N} \\ \text{Num} \\ \text{Adj} \\ \text{Num} \\ \text{Pron} \\ \text{Vh} \end{array} \right\} +\text{-дан} + \left\{ \begin{array}{l} \text{N} \\ \text{V} \\ \text{Adj} \\ \text{Vh} \\ \text{Adv} \end{array} \right.$	$\left. \begin{array}{l} \text{N} \\ \text{Adj} \\ \text{Num} \\ \text{Pron} \\ \text{Vh} \end{array} \right\} + (\text{кел.афф}) + \text{кўмакчи} + \left\{ \begin{array}{l} \text{N} \\ \text{V} \end{array} \right.$
---	---

Қуйида бошқарувли бирикмалар моделларига мисоллар келтирилди:

№	Бошқарувли бирикмалар модели	мисоллар
1	Ncom-ни+ V	аризани ёзмоқ, мурожаатни кўриб чиқмоқ
2	Adjни + V	қизилини танламоқ
3	Numни + V	учни тўлдирмоқ, иккисини бирлаштирмоқ
4	Pronни(дир) + V	хаммани соғинмоқ, кимнидир кўрмоқ
5	Vhни + V (Vh)	ўқишни бошламоқ, ёзишни ўргатиш
6	Vhга + V	теришга кетмоқ
7	Ncom-га+ V	мақсадга ўтмоқ
8	Adjга + V	баландга кўтарилмоқ, оққа кўчирмоқ
9	Numга + V	бирга кирмоқ, бешга ўтмоқ
10	Pron(лар)га(дир) + V (Vh)	хаммадан сўрамоқ, кимгадир кўмаклашиш
11	Vhга + V	таълим олишга кўмаклашиш
12	Ncom-да+ V	уйда ўтирмоқ, идорада ишламоқ
13	Numда + V	саккизда келмоқ, тўртда кетмоқ
14	Pronда + V	унда тажриба ўтказмоқ
15	Vhда + V	(мутолаа) қилишда фаоллашмоқ
16	Ncom-дан+ Vh	фикрдан воз кечиш, мозийга қайтиш
17	Adjдан + V (Vh)	сарифидан олмоқ
18	Numдан + V	иккисидан қолишмасдан
19	Pronдан(дир) + V	кимдандир сўраш
20	Vhdан + V	ҳаракатланишдан тўхтамоқ
21	Vr + V	шошиб сўзламоқ
22	Ncom(эғалик афф.) + Кўм.(кўмакчи)+Ncom	халқ билан мулоқот

23	Ncom(эгалик афф.)+Кўм. (кўмакчи)+ Ncom	онам учун совға
24	Pron + Кўм.(кўмакчи) + Adj	мен учун азиз
25	Ncom-дан + Adj	отдан баланд, ойдан гўзал, темирдан қаттиқ
26	Ндан-да+ Adj	кундан-да гўзал
27	Adj-дан + Adj	шириндан ширин
28	Numдан + Num	бешдан икки
29	Numга + Num	учга уч
30	Ncom-қа + Ncom	халққа мурожаат
31	Ncom-га +	оқимга қарши
32	Vh + Ncom	нураган уй
33	Ncom-дек + Ncom	фариштадек одам
34	Pron-дан + Adv	ҳаммадан ўзгача
35	Pron-дан + Adj	ҳаммадан чиройли
ва ҳ.к.		

Сўз бирикмалари моделлари бизга 2 мақсадда керак:

- 1) N-граммалари сўзлар бирикмасини таҳлил қилиш;
- 2) Омонимшакларнинг лингво-таҳлил технологиясини амалга ошириш.

Умуман, омоним-таҳлил технологияси муайян омоним сўзга бошқа сўзнинг маълум грамматик кўрсаткич орқали боғланиши билан тушунтирилади ва бу жараён таҳлил имконияти ҳамда таҳлилнинг аниқлигини оширади.

Мазкур лингвистик усул барча лингвистик дастурлар замирида ишлаши билан аҳамиятли ҳисобланади.

Технологиянинг татбиқ қилиниши натижасида ўзбек тилидаги матнлар устида ишлашга мўлжалланган, жумладан, машина таржимаси, лингвистик корпус, матнларни қайта ишлаш системалари, автоматик таҳрир ва таҳлил дастурлари каби лингвистик дастурий таъминотларда омонимларни таҳлил қилиш муаммоси бартараф этилади.

Умуман, омоним-таҳлил технологияси муайян омоним сўзга бошқа сўзнинг маълум грамматик кўрсаткич орқали боғланиши билан тушунтирилади ва бу жараён таҳлил имконияти ҳамда таҳлилнинг аниқлигини оширади.

Хулоса қилиб айтганда, дунё компьютер лингвистикасида омонимликни бартараф этиш усуллари ўрганилганида, бу хусусдаги тажрибадан фойдаланиб ўзбекча матнлардаги сўзшаклларнинг тегишли текшириш формуласи яратилди. Омонимликни бартараф этиш учун ҳар бир сўзни "таснифлаш" керак, яъни уни лемма - гап бўлаги ва морфологик хусусиятлар мажмуи билан таққослаш мумкин, улар қулайлик учун бир тегга қўшилади. Барча мумкин бўлган тегларни ўрганиш учун морфологик лугатдаги сўзларга тегишли ҳаволаларни топиш ёки MyStem 12 каби морфологик анализаторни ишлатиш етарли бўлиб, у сўз тегларини топишда ёрдам беради. Шундан сўнг бир нечта теглар орасидан фақат тегишли тегни танлаш керак бўлади.

Омоним сўзшаклларни таҳлил қилишда қўлланилган оптимал лингвистик технология матнларни таҳрир ва таҳлил қилиш, машина таржимаси, матнларни қайта ишлаш жараёнларида муҳим омил бўлади.

Фойдаланилган адабиётлар:

1. Baum, L. E.; Sell, G. R. Growth transformations for functions on manifolds. Pacific Journal of Mathematics. 27 (2) 1968. - P. 211-227.; https://en.wikipedia.org/wiki/Hidden_Markov_model.
2. Brill E. A simple rule-based part of speech tagger //Proceedings of ANLC. 1992, - С. 154.
3. Brill E. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging // Computational Linguistics. Vol. 21. 1995. - P. 543-565. <http://acl.ldc.upenn.edu/J/J95/J95-4004.pdf>
4. Hearst M.A. Noun homograph disambiguation using local context in large text corpora // Proceedings of the 7th conference, UW Centre for the New OED & Text Research Using Corpora, 1991.
5. Зеленков Ю.Г., Сегалович И.В., Титов В.А. Вероятностная модель снятия морфологической омонимии на основе нормализующих подстановок и позиций соседних слов (Электрон ресурс). http://www.dialog-21.ru/media/2444/zelenkov_segalovich.pdf
6. Клышинский Э.С., Рысаков С.В. Статистические методы снятия омонимии // "Новые информационные технологии в автоматизированных системах". 2015. - С. 556-557.
7. Кобрицов Б.П. Методы снятия семантической неоднозначности. НТИ, Сер.2, Вып. 3, 2004.
8. Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. Автореф... канд. филол. наук. Москва: РГГУ, 2004.
9. Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Снятие лексико-семантической омонимии в новостных и газетножурнальных текстах: поверхностные фильтры и статистическая оценка (Электрон ресурс). http://elar.urfu.ru/bitstream/10995/1388/1/ИМАТ_2005_03.pdf.
10. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. - СПб, 2006. - С.26.
11. Очиқ рус миллий корпуси сайти - <http://opencorpora.org>.
12. Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. - Тошкент: Ўқитувчи, 1984. - Б.5.
13. Ризаев С. Ўзбек тилшунослигида лингвостатистика асослари. - Тошкент: Фан, 2006. - Б. 18.
14. Рысаков С.В. Методы борьбы с омонимией. <http://samag.ru/archive/article/3059>.
15. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан. бўйича фалсафа докт. (PhD) дисс. автореф. - Қарши, 2018.
16. <http://tech.yandex.ru/mystem> - MyStem морфологик таҳлил дастури сайти.
17. <http://www.academia.edu/15517740/>
18. www.dialog-21.ru/media/2138/zakharov.pdf Захаров В.П. Корпуса русского языка.