

Morphologic Analyzer – as a Mathematician Model

E'zoza F. Sharipova

Student

Gulistan State Pedagogical University

Gulistan, Syrdarya Region

Uzbekistan

Abstract

The article provides a brief overview of corpus linguistics issues from introduction to conclusion. That is, putting any social and humanitarian sciences into mathematical models, in short, modeling or digitalization will form an innovative environment in the world of science.

Keywords: morphological analyzer, automatic analysis, graphematic analysis, morphological analysis, syntactic analysis, semantic analysis, lemma, stem, tag, prosodic tag, anaphoric tag, semantic tag, morphological tagging, syntactic tag

Морфологик анализатор – математик модел сифатида

Шарипова Эъзога Фазлиддиновна

Талаба

Гулистон давлат педагогика институти

Гулистон ш., Сирдарё вилояти

Ўзбекистон Республикаси

Аннотация

Ушбу мақолада корпус лингвистикасига доир масалаларнинг муқаддимасидан то хотимасигача бўлган ҳолатнинг қисқача мазмуни ёритиб берилган. Яъни ҳар қандай ижтимоий-гуманитар фанларни математик моделларга солиш, қисқачароқ қилганда, қолиплаш ёки рақамлаштириш фан оламида инновацион муҳитни шакллантириб беради.

Калим сўзлар: морфологик анализатор, автоматик таҳлил, графематик таҳлил, морфологик таҳлил, синтактик таҳлил, семантик таҳлил, лемма, стем, тег, просодик тег, анафорик тег, семантик тег, морфологик теглаш, синтактик тег

1.0. Кириш

Морфологик анализатор – луғат(лексикон)даги сўзшакл ва алоҳида сўзларни қиёслашга мўлжалланган, сўзнинг грамматик хусусиятини ифодалайдиган алгоритмлар тўплами (электрон ресурс). Синтактик парсер матнни "таниш" функциясиз ҳам ишлай олса-да, таҳлил жараёнида морфологик анализатор таҳлил натижасини фаол қўллайди. Матннинг грамматик ахборот берувчи разметкаси синтактик таҳлил қоидаларини тузишга ёрдам бериши айtilган. Масалан, от ва сифат орасидаги муносабатни аниқлаш қоидасини олиш мумкин. Рус грамматикаси қоидаларидан келиб чиқиб айтиш мумкинки, отдан чап томонда турувчи сифат от билан бир хил келишик, сон ва родда бўлса, у билан алоқага киришади: на крутом (предложный, ед., муж.) берегу (покатив; ед., муж.). Ушбу қоида от ва сифат боғланишининг ягона асоси ҳисобланмайди, аммо у нисбатан кўпроқ ҳолатларда шундай боғланишларни ҳосил қилади (Шарипов, 2023; Хамроева, 2020).

Тадқиқотнинг объекти сифатида корпус лингвистикасида морфологик анализаторлар жорий этилган жаҳон ва ўзбек илмий тилшунослигидаги илмий асарлар компьютер лингвистикаси билан боғлиқ тарзда ёритилган бўлиб, жумладан, жаҳон, рус ва ўзбек тилшунослигининг грамматикасига оид илмий қарашлар ҳам асос қилиб олинди. Бундан ташқари мақолада жаҳон ва ўзбек тилшунос олимларининг илмий тадқиқотларидан, яъни компьютер лингвистикаси, қолаверса, миллий корпус лингвистикаси билан боғлиқ мунозарали масалаларга ойдинлик киритганлиги тўғрисидаги маълумотлар ўрин эгаллаган. Тадқиқот мавзусини ёритишда таснифлаш, тавсифлаш, тарихий-қиёсий ва функционал таҳлил усулларидан фойдаланилган.

2.0. Автоматик таҳлил

Матннинг автоматик таҳлили бир неча мураккаб операциядан иборат: компьютер табиий тилдаги матн таҳлилини белгиланган алгоритм асосида амалга оширадики, автоматик таҳлил давомида компьютер матн ҳақида ўз тилида лексик-морфологик, синтактик, семантик тасаввур ҳосил қилади. Матннинг автоматик анализ ва синтез жараёни компьютер лингвистикасининг асосий вазифаларидан бири саналади. Бу вазифа сунъий интеллект учун лингвистик моделлар яратиш ва инсониятнинг амалий эҳтиёжидан (масалан, автоматик машина таржимаси) келиб чиқади (Щипицина, 2013).

Матннинг автоматик таҳлили бир неча босқичдан иборат:

- 1) графематик таҳлил: сўз, гап, абзац ва бошқа матний элемент (масалан, газета матни бўлаги) чегарасини аниқлаш;
- 2) морфологик таҳлил: матнда қўлланган сўзшаклнинг бош шаклини ҳамда бу сўзнинг морфологик белгиларини аниқлаш;
- 3) синтактик таҳлил: матндаги гапнинг грамматик структурасини аниқлаш;
- 4) семантик таҳлил: ибораларнинг маъносини фарқлаш (Автоматическая обработка текстов на естественном языке и компьютерная лингвистика, 2011).

Юқорида кўрсатилган таҳлил турларидан бири графематик таҳлил токенизация (инг. token. – матндаги сўз, ибора ёки бирор аҳамиятли элемент)га тенг бўлиши мумкин. Матний элементлар чегараси аниқловчиси сифатида шаклий-сигнал сўзлар орасида пробел (ажратма) бўлса, гап ва унинг қисмлари чегараси бош ҳарф ва тиниш белгилар, матннинг мазмуний қисмлари абзац билан фарқланади (Щипицина, 2013). Аммо сўз чегарасини аниқлашда доим ҳам формал метод қўл келмайди. Масалан, хитой тилида сўзнинг шаклий чегараси мавжуд эмас. Ҳатто Европа тилларининг аксариятида пробел (ажратма) билан ажратилган бир неча сўзларнинг битта лексемага тенг келиш ҳолати ҳам бу метод билан графематик таҳлил натижа бермаслигини кўрсатади.

Иккинчи таҳлилга назар ташлайдиган бўлсак, яъни матннинг морфологик таҳлилида унда қўлланган сўзшаклнинг бош шакли – лемма аниқланади ҳамда сўзнинг шу контекстдаги грамматик маънолари ажратилади. Масалан, от учун туркум, шахс, сон, келишик, феъл учун шахс, сон, замон ва ҳок... Матнда ишлатилган ҳар бир сўз, сўзшакл ёки сўз қўллаш ҳолати дейилади. Флектив тилларда морфологик таҳлилдаги ўзак (асос) анъанавий грамматикадаги асосга тўғри келмаслиги мумкин. Морфологик таҳлилда асосни аниқлаш стемнинг дейилади. Стем – бу сўзшаклнинг синтактик шакл ясовчиларсиз қисми, флексия ҳодисасига учраган айрим туркий ўзакларнинг грамматик асоси билан морфоанализ жараёнидаги асоси – стем мос келмай қолиш ҳолатлари учраб туради. Масалан, сон ўзагидан ясалган сана феълнинг ўзаги сон, аммо бу сўзшаклнинг стем сан. Бу ерда ўзакдан -а сўз ясовчи ажратиб олинади ҳамда сан стем бўлиб қолади. Лекин агглютинатив тилларда, жумладан, ўзбек тилида одатда, лемма стемга тенг келади ҳамда флексия ҳолати морфоанализ жараёнига кўп ҳалақит бермайди (Хамроева, 2020).

Морфоанализда фақат лемматизация жараёнининг ўзи камлик қилади, бунда стемнингнинг ҳам ўз ўрни бор. Хусусан, интернетда қидирув сифатини яхшилашда стем асосий омил бўлади, яъни қидирув лемма асосида эмас, стем асосида амалга оширилади (Толдова и др., 2011). Демак, морфологик таҳлилда компьютернинг сўзшаклни тушуниши билан иш битмайди, яъни белгилар кетма-кетлигининг компьютер томонидан фарқланиши етарли бўлмайди. Сўзшаклнинг парадигмадаги барча аъзоларидан фарқини топиш жараёни билан таҳлил тугамайди. Балки сўзшаклнинг туркумга ҳослиги (тагнинг) ҳамда морфологик белгиларини топиш ҳам асосий вазифа, грамматик маъносини билдирувчи қўшимчалар (компьютер моделлари) муайян лексеманинг шакл ясовчилари ҳамда унинг парадигма аъзоси сифатида шакллантирилса, мақсадга мувофиқ. Сўзлардаги лексик-морфологик омонимия ҳолатини аниқлаш учун сўзшаклга морфологик белгиларни тўғри бириктириши учун гапнинг синтактик ҳамда семантик хусусиятига таянилади. Хуллас, автоматик ҳамда инсон омили билан амалга оширилиши назарда тутилган. Бу каби жараёнларнинг автоматик тегланишида қоидалар модуллари ишга туширилади.

Кўриниб турибдики, бундай қоидалар мажмуи, яъни гапда сўз туркумлари ва гап бўлақларининг тартибини аниқлаш орқали омонимияни фарқлаш имконияти пайдо бўлади.

3.0. Морфологик анализатор

Ш.Хамроевнинг монографиясида морфологик анализатор тузишнинг уч усули санаб ўтилади:

- 1) луғатга асосланган анализатор;
- 2) луғатсиз грамматик қонуниятлар мажмуига асосланган анализатор;
- 3) луғат ва грамматик қоидалар базасига асосланган анализатор (Хамроева, 2020).

Бунда Олтой оиласига кирувчи тиллар орасида сингармонизм ҳодисаси нисбатан кенг тарқалган қирғиз тили қоидалари ҳамда далилларига таянганлигига ва тўрт босқичда амалга оширилишига гувоҳ бўламиз:

- 1) кириш матни сўзларини грамматик шаклларга ажратиш;
- 2) сўзни лемматизациялаш, яъни асослар луғатидан, сўзнинг лексема шаклини топиш, аниқлаш;
- 3) синтактик шакл ясовчи қўшимчалар занжирини аниқ қисмларга ажратиш;
- 4) ҳар бир аффикснинг морфологик белгисини аниқлаш.

Тадқиқотчи томонидан морфологик анализаторнинг моҳиятини тушунтириш учун ўзбек ва қирғиз тили далилларини қиёсланганлигини кўрамиз. Масалан, балдар (болалар) сўзшакли таҳлил қилинаётганда, морфологик анализатор бу шакл -лар қўплик қўшимчаси қўшилган бола асосидан ҳосил бўлганлигини аниқлаши керак. Асосга -лар қўшимчаси қўшилиши натижасида қирғиз тилидаги сингармонизм ҳодисаси бала асосидаги охириги ҳарфнинг тушиб қолишига олиб келган. Бундай ҳолатда анализатор сўзни леммалаши, қўшимчадан олдинги асосни автоматик аниқлаши лозим. Бунда асослар луғати қидируви ишга туширилади. Китобим сўзшаклини таҳлил қилишда морфологик анализатор унинг китоб сўзига -им эгалик шакли қўшилганлигини аниқлаши лозим: китоб асос эканлиги аниқланади. Қирғиз тилида китебим эгалик шаклидаги сўз китеп ўзагига эгалик қўшимчаси қўшилганда -им қўшимчаси таъсирида жарангсиз п жарангли б товушига айланади. Бундай ҳолатда асослар луғати қўшимча қоидалар билан таъминланади. Ўзбек тилида фонетик ўзгаришларнинг катта миқдорни ташкил этмаганлиги морфологик таҳлил жараёнида алгоритмлар босқичини қисқартириши мумкин. Аммо фонетик ўзгариш учрайдиган ҳолатда худди қирғиз тили каби ўзгаришга учраган сўзнинг ўзагини аниқлаш жараёни 2–3-босқичли алгоритмда амалга оширилади (Хамроева, 2020).

4.0. Хулоса

Бундан шундай хулосага келинадики, ўзбек тили морфологик анализатори тузишнинг лингвистик таъминотини ишлаб чиқиш тамойиллари орасида мана шу қоидалар блоки – лингвистик модулни шакллантиришнинг асосий вазифаси саналади.

Адабиётлар рўйхати

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие. – Москва: МИЭМ, 2011. – С. 106–107.
2. Толдова С. Ю., Бонч-Осмоловская А. А. Автоматический морфологический анализ // Фонд знаний «Ломоносов». – Москва, 2011.
3. Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. – Тошкент, 2020. – Б. 50.
4. Щипицина Л. Ю. Информационные технологии в лингвистике: учебное пособие. – Москва: Наука, 2013. – С. 43.
5. [E-resource] URL: https://www.solarix.ru/for_developers/docs/morphology_analyzer.html
6. Sharipov F. Hozirgi o'zbek adabiy tili. – Toshkent: BOOKMANY PRINT, 2023. – 100 b.

Тақризчи:

Ф. Г. Шарипов

Гулистон давлат университети профессори, филология фанлари доктори

Reviewer:

F. G. Sharipov

DSc in Philological Sciences, Professor of Gulistan State University