

ФОРМАЛ ГРАММАТИКА НАЗАРИЯСИ

Жураева Нилуфар Вайитовна,

Чирчик давлат педагогика институти,
информатика кафедраси доценти

Хасанова Суманбар Хамроқуловна,

Чирчик давлат педагогика институти,
информатика кафедраси ўқитувчиси

Аннотация: Мазкур мақолада сунъий тиллар, уларнинг классификацияси ва ишлаш принциплари ҳақида маълумот берилган. Шу принцип асосида табиий тил, яъни ўзбек тили мисолида, формал грамматикаси келтирилган. Олинган натижалар ўзбек тили ва адабиёти йўналиши талабалари учун компьютер лингвистикаси дарсларини ўтишда қўлланилиши мумкин.

Калит сўзлар: табиий тил, компилятор, транслятор, интерпритатор, формал грамматика, синтаксис, лексема.

Компьютерлар билан мулоқот қилиш учун одамларнинг табиий мулоқот тилларидан фарқ қиладиган махсус тиллар ишлаб чиқилади, улар сунъий тиллар дейилади. Сунъий тиллар, бир томондан, одамлар учун қулай ва тушунарли бўлиши керак, бошқа томондан, қурилмалар томонидан “идрок” этилиши керак. Ушбу талабларни битта тилда бирлаштириш қийин вазифа, шунинг учун инсон тушунадиган матнларни қурилма тушунадиган тилга алмаштириш воситалари қўлланилади. Ушбу воситалар трансляторлар яъни таржимонлар деб номланади.

Таржимонни *талқин* қилиш ёки *тузиш* мумкин. Биринчисида у кириш тили учун транслятор, иккинчисида эса компилятор деб номланади. Трансляторлар, *интерпритатор* ва *компилятор* турларига ажралади.

Интерпритатор кириш тилининг гапларини кетма - кет ўқийди ва таҳлили қилиб дарҳол бажаради. Компилятор (тузувчи) тилнинг жумлаларини бажармайди, балки кейинчалик натижани олиш учун дастур тузади. Компиляторга инсон тилидаги матн берилади, натижавий матн эса қурилма тушунадиган кўринишда тузилади.

Компиляторнинг иши бир неча босқичлардан иборат бўлиб, улар кетма-кет бажарилиши мумкин. Компиляторнинг ишининг *биринчи босқичи* лексик таҳлил дейилади, дастури эса - лексик таҳрирчи (ЛТ). Лексик таҳлилчига

кириш тилидан белгилар кетма-кетлиги киритилади. ЛТ ушбу кетма-кетликда лексик бирликлар деб аталадиган энг содда конструкцияларни ажратиб олади. Лексик бирликларга мисол қилиб, идентификаторлар, рақамлар, операцион белгилар, хизмат сўзлари ва бошқаларни олиш мумкин. ЛТ матндаги лексик бирликларни лексема деб номланган, уларнинг ички кўринишлари билан алмаштиради. Лексема таркибида лексик бирликлар синфи ва уларнинг таърифи ҳақида маълумот бўлиши мумкин.

Компилятор ишининг иккинчи босқичида синтактик тахрирловчига (СТ) лексемалар кетма кетлиги киритилади, бу ўз навбатида, ҳаракатлар белигиси кетма кетлигидан иборат оралик кодга алмаштирилади.

Компилятор ишининг учинчи босқичида чиқувчи матн тузиш амалга оширилади. Бу жараёни амалга оширадиган дастур чиқувчи матн генератори(Г) деб номланади. Генератор ўнга кирувчи ҳар бир ҳаракат белгисига, чиқувчи тилдаги бир ёки бир неча буйруқларни мослайди. Чиқувчи тил сифатида қурилма буйруқлари, ассемблер буйруқлари ёки бошқа бир тилнинг операторлари ишлатиши мумкин.

Компиляторни тузиш учун кириш ва чиқиш тилларининг аниқ ва бир қийматли таърифи талаб қилинади. Бундай талаб тилнинг мақбул конструкцияларини қуриш қоидаларини аниқлашда керак бўлиши мумкин. Ушбу қоидалар тўплами тил синтаксиси деб аталади. Бундан ташқари, топшириқ ўз ичига ҳар бир конструкция маъноси ва мақсад тавсифидан иборат бўлиши керак. Ушбу тавсиф тилнинг семантикаси деб номланади.

Киритилган матнни ўрнатилган қоидалар асосида тузилган белгилар кетма-кетлиги деб қараш мумкин. Формал тил ёки грамматика деб, матнларни ифодалашда ишлатилган белгилар кетма-кетлигидан иборат математик моделга айтилади.

Қуйида формал грамматика қоидалари ва талаблари асосида табиий тил Ўзбек тили мисолида формал грамматикасини бериш ҳаракат қилинган.

Формал грамматика Ўзбек тили мисолида

Тилнинг асосий элементлар бўлган тушунчаларга формал грамматика нуқтаи назарида таъриф бериб ўтамыз:

Сўз - маънога эга бўлган, муаян тилнинг луғатида мавжуд белгилар кетма кетлиги. Тилнинг лексикаси деб, тилда жоиз бўлган сўзлар тўпламига айтилади.

Гап - сўзларнинг мураккаб бўлган конструкциялар бирлашмасига айтилади. Тил бу гаплар тўпламидир. Гаплар синтактик қоидалар асосида

сўзлар ва содда гаплардан тузилади.

Синтаксис деб, тўғри тузилган гапларнинг тавсифи тушунилади.

Алфавит, лексика ва синтаксис тилдаги жоиз конструкциялар тўпламини ва конструкциялар орасидаги ички муносабатларни тўлиқ аниқлайди.

Синтаксис қоидалар тўплами тилнинг грамматикасини ташкил этади. Синтаксис қоидаси тўғри гапларни ҳосил қилиш процедурасини ёки гапнинг “тўғрилигини” (яъни гапни мазкур тилга тегишлилигини) аниқлаш процедурасини ифодаши мумкин. Биринчи ҳолатда, грамматика *генератив*, иккинчи - *аниқланувчи* грамматика дейилади.

Мисол: Айтайлик G_0 грамматика қўйидаги қоидалар тўплами билан аниқланган.

$\langle \text{гап} \rangle \rightarrow \langle \text{эга} \rangle \langle \text{кесим} \rangle$

$\langle \text{эга} \rangle \rightarrow \langle \text{от} \rangle$

$\langle \text{эга} \rangle \rightarrow \langle \text{олмош} \rangle$

$\langle \text{от} \rangle \rightarrow \text{олма}$

$\langle \text{олмош} \rangle \rightarrow \text{у}$

$\langle \text{кесим} \rangle \rightarrow \langle \text{феъл шакли} \rangle$

$\langle \text{феълли шакли} \rangle \rightarrow \text{тушди}$

\rightarrow белги «бу... дегани» деб ўқилади. У ҳар бир қоидани 2 қисмга ажратади: *ўнг ва чап*.

Маълум қисқартиришларни киритилса, G_0 грамматикани қисқача қўйидагича ёзиш мумкин:

$\langle \Gamma \rangle \rightarrow \langle \text{Э} \rangle \langle \text{К} \rangle$

$\langle \text{Э} \rangle \rightarrow \langle \text{ОТ} \rangle$

$\langle \text{Э} \rangle \rightarrow \langle \text{ОЛМ} \rangle$

$\langle \text{ОТ} \rangle \rightarrow \text{олма}$

$\langle \text{ОЛМ} \rangle \rightarrow \text{у}$

$\langle \text{К} \rangle \rightarrow \langle \text{ФШ} \rangle$

$\langle \text{ФШ} \rangle \rightarrow \text{тушди}$

Ихтиёрий грамматиканинг қоидалар тўпламини P деб белгилаймиз. Қоида таркибидаги белгилар V луғатни ташкил этади.

$V = \{ \langle \Gamma \rangle, \langle \text{Э} \rangle, \langle \text{К} \rangle, \langle \text{ОТ} \rangle, \langle \text{ОЛМ} \rangle, \langle \text{ФШ} \rangle, \text{олма}, \text{у}, \text{тушди} \}$

Одатда, луғат элементлари катта харфлар билан белгиланади. V луғатдаги барча чекли белгилар (қаторлар) кетма кетлиги тўплами ҳамда бўш қаторни V^* деб белгилаймиз. V^* тўплам элементларини кичик харфлар билан белгиланади. V тўплам иккита тўплам остига бўлинади: 1) P

қоиданинг фақат ўнг томонига кирувчи T символлар қисм тўплами ($T = \{ \text{олма, у, тушди} \}$); 2) N қисм тўплам $= V \setminus T$ ($N = \{ \langle \Gamma \rangle, \langle \Xi \rangle, \langle K \rangle, \langle OT \rangle, \langle OLM \rangle, \langle \Phi \Pi \rangle \}$).

T қисм тўпламнинг символлари *терминалли* ёки *терминал* (чекли) деб номланади. N қисм тўплам элементлари эса *терминал бўлмаган* ёки *ўзгарувчи* дейилади. Грамматикада, терминал бўлмаган символлар ажралиб туради. Уни грамматиканинг бошланғич (аксиома) белгиси дейилади ва у A ёки S ($A = \{ \langle \Gamma \rangle \}$). Бу белги аниқланувчи тилни ифодалайди.

$G = \{ V, T, P, A \}$ ни аниқлайдиган G грамматикани, P қоидалар тўплам элементлари қуйидаги кўринишга эга: $x \rightarrow y$, ($x \neq y$, $x \in (V \setminus T)^*$, $y \in V^*$), генератив дейилади, $x \rightarrow y$ кўринишидаги қоида генератив қоидалар ёки ўрнига қуйиш қоидаси дейилади.

Формал тиллар

G генератив грамматика $L(G)$ тилни ҳосил қилади. Бу қандай бўлишини кўриш учун қўшимча таърифларни киритамиз.

$x \in V^*$ қатор тўғридан-тўғри $y \in V^*$ ($x \Rightarrow y$ деб белгилаймиз) қаторни ҳосил қилади агар: $x = pUq$, $y = p z q$. Бу ерда $p \in V^*$, $q \in V^*$, ва $U \rightarrow z \in P$ қоида мавжуд. Бошқача қилиб айтганда, y – бу x символнинг тўғридан-тўғри ҳосиласи (келиб чиқиши) бўлади, агар қоида бўйича у ни x га терминал бўлмаган U белгисини z қатори билан алмаштириш орқали олиниши мумкин бўлса. Мисол: G_0 грамматикасида U тушди қатори, $\langle OLM \rangle \rightarrow U$ қоидани қўллаб олинган $\langle OLM \rangle \rightarrow$ тушди қаторининг ҳосиласи.

$x \in V^*$ қатор $y \in V^*$ қаторни ҳосил қилади ($x^* \Rightarrow y$ каби белгиланади), агар шундай $x = x_0, x_1, \dots, x_n = y$ қатор мавжуд бўлсаки, $x_i \Rightarrow x_{i+1}$, $i = 0, 1, \dots, n - 1$ ўринли бўлса. Бундай кетма-кетлик қатори келтириб чиқариш дейилади. Мисол. G_0 грамматикада U тушди қатори $\langle \Gamma \rangle$ қаторнинг ҳосиласи:

$$\langle \Gamma \rangle \Rightarrow \langle \Xi \rangle \langle K \rangle \Rightarrow \langle OLM \rangle \langle K \rangle \Rightarrow U \langle K \rangle \Rightarrow U \langle \Phi \Pi \rangle \Rightarrow U \text{ тушди}$$

A белгиси томонидан ҳосил қилинган ва фақат терминал белгиларидан иборат бўлган сатрлар, $L(G)$ тилини ташкил қилади (улар тил гаплари деб аталади). Мисол. $L(G_0) = \{ \text{олма тушди, у тушди} \}$

$L(G_0)$ тилида 2 тўғри гап бор, $L(G_1)$ тилида уларнинг сони чексиз. Бу G_1 грамматикасининг рекурсивлиги натижасидир. Ҳар бир гапга камида битта келтириб чиқариш қоидаси тўғри келади. Бироқ, бир нечта келтириб чиқариш келтириб чиқариш битта гапга мос келиши мумкин. Мисол. G_0 грамматикада U тушди қатори $\langle \Gamma \rangle$ қаторининг ҳосиласи:

$$\langle \Gamma \rangle \Rightarrow \langle \Xi \rangle \langle K \rangle \Rightarrow \langle \Xi \rangle \langle \Phi \Pi \rangle \Rightarrow \langle \Xi \rangle \text{ тушди} \Rightarrow \langle OLM \rangle \text{ тушди} \Rightarrow U$$

тушди

Келтириб чиқариш *чан томонлама* дейилади, агар ҳар бир қадамда *чан* ўзгаруви алмашса ва келтириб чиқариш *ўнг томонлама* дейилади, агар ўнг ўзгарувчи алмашса.

Чан томонлама келтириб чиқариш:

$\langle \Gamma \rangle \Rightarrow \langle \mathcal{E} \rangle \langle K \rangle \Rightarrow \langle \text{ОЛМ} \rangle \langle K \rangle \Rightarrow \mathcal{U} \langle K \rangle \Rightarrow \mathcal{U} \langle \Phi \text{Ш} \rangle \Rightarrow \mathcal{U}$ тушди

Ўнг томонлама келтириб чиқариш:

$\langle \Gamma \rangle \Rightarrow \langle \mathcal{E} \rangle \langle K \rangle \Rightarrow \langle \mathcal{E} \rangle \langle \Phi \text{Ш} \rangle \Rightarrow \langle \mathcal{E} \rangle$ тушди $\Rightarrow \langle \text{ОЛМ} \rangle$ тушди $\Rightarrow \mathcal{U}$

тушди

Келтириб чиқаришни синтактик дарахт деб тасаввур қилиш мумкин, ёки бошқача қилиб айтганда *келтириб чиқариш дарахти* ёки *таҳлил дарахти*.

Агар ҳеч бўлмаганда битта гап биттадан кўп синтактик дарахтга эга бўлса, *ноаниқ (бирқиймали эмас) грамматика* деб аталади.

Танлаш масаласига тескари масала - *таҳлил масаласи*. Генеративга тескари қаторларни ўзгартириш *қаторлар редукцияси* (мураккабдан соддага ўтиш) дейилади.

$u \in V^*$ қатор тўғридан тўғри $x \in V^*$ қаторга келади, агар x тўғридан тўғри u ни ҳосил қилса. $u \in V^*$ қатор $x \in V^*$ қаторга келади, агар x u ни ҳосил қилса.

Синтактик таҳлилининг асосий масаласи, тилдаги берилган гапнинг таҳлилини топишдир. Агар таҳлили мавжуд бўлса, у ҳолда гап синтасис жихатидан тўғри. Таҳлил, уни структурасини беради (синтактик дарахтини). Таҳлил масаласини ҳал қилувчи алгоритм англовчи дейилади.

Фойдаланилган адабиётлар:

1. Булыгина Т.В., Крылов С.А. Модель // Языкознание. Большой энцикло- педический словарь / Гл. ред. В.Н. Ярцева. – 2-е изд. – М.: Большая Российская энциклопедия, 2000. – стр.304-305
2. Juraeva N. V., The formal model of the grammar of the uzbek language . European Journal of Research and Reflection in Educational Sciences// Vol.8 No. 4,2020 Part II ISSN 2056-5852., pp. 129-133.
3. Juraeva N.V., Sultanov R.O., Abdullaeva S.A., Rakhimjonova V. A. Systetization uzbek language word combinations. Международный журнал «Наука и мир». Vol.II. No. 6(82), 2020. ISSN 2308-4804., стр. 65-68.
4. Жураева Н. В., Жураева У. В. Формальные модели членов



предложения грамматики узбекского языка в рамках официально-делового
стиля. Теория и методика профессионального образования. Universum.
Психология и образование.12(78). 2020.