

# Унификационни граматики (УГ)

## I. Увод

Граматиките са абстрактни машини, които:

1. Определят дадено множество от изречения като правилно построени изречения.
2. Дефинират съответствие (изображение) между изреченията и тяхната лингвистична структура.

Има два подхода към граматиките:

– логически подход: Граматиката се разглежда като формална система. Речникът е множеството от аксиоми. Правилността на дадено изречение и фактът, че то има някаква лингвистична структура, са теореми, които могат да се изведат от аксиомите чрез правилата на граматиката. Проблемът за анализа се разглежда като задача за определяне дали дадена формула е следствие от аксиомите. DCG се базират на този подход.

– подход на ограниченията: Граматиката се разглежда като множество от ограничения, като правилно-построена лингвистична структура е всяка структура, която удовлетворява тези ограничения. Едно изречение е правилно построено, когато неговата лингвистична структура е правилно построена структура. УГ се базират на този подход.

Обикновено една граматика може да се разглежда от гледна точка и на двата подхода.

## II. Информационни структури на УГ

УГ се състоят от два компонента:

1. Безконтекстни (фразово-структурни) правила, наричани още гръбнакът на граматиката.

2. Информационни структури, които:

– описват ограниченията

– описват лингвистичната структура на изречението

Тези информационни структури се наричат *атрибутни структури* (feature-value structures, attribute-value structures).

Всяка атрибутна структура (АС) е множество от елементи от вида:

**<атрибут : стойност>.**

Атрибутите характеризират обекта, който описваме с тази АС. Описанията обикновено са непълни – вземат се само тези атрибути (свойства,

характеристики) на обекта, които ни интересуват. Напр. за синтаксиса – лице (person), число (number), време (tense), ...

Атрибутите са уникални за всяка структура, т.е. не може да има два елемента с едно и също име на атрибут. (В една АС не може да има две наредени двойки с еднакъв първи елемент.)

Стойностите на атрибутите могат да бъдат:

- константи (Дефиниционната област на стойностите се определя от обекта (езика). Например, за число – {‘ед.ч.’, ‘мн.ч.’} или number – {s,p}.);
- променливи;
- атрибутни структури.

АС може да се представи и чрез ориентиран граф без цикъл (DAG - Directed Acyclic Graph) с ребра - имената на атрибутите и възли - стойностите на атрибутите

АС е частично определено изображение от атрибутите към техните стойности.

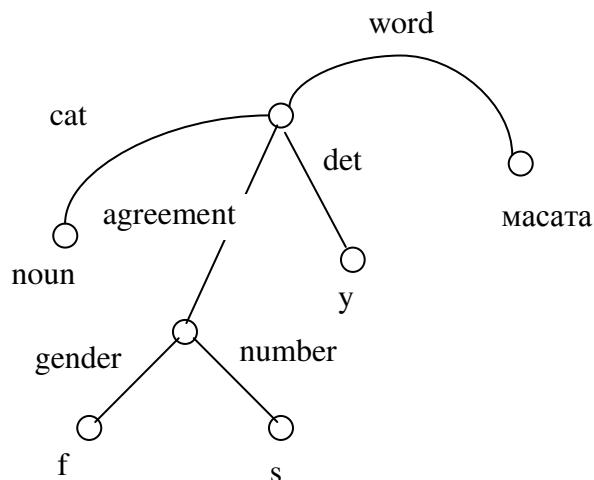
Пример за описание на думи чрез АС (речникът може да се състои от такива описания):

<word: "масата"> <cat: noun> <gender: f> <number: s> <det: y>

В друг запис:

$\left[ \begin{array}{ll} \text{word} & : \text{масата} \\ \text{cat} & : \text{noun} \\ \text{gender} & : f \\ \text{number} & : s \\ \text{det} & : y \end{array} \right]$	ИЛИ	$\left[ \begin{array}{ll} \text{word} & : \text{масата} \\ \text{cat} & : \text{noun} \\ \text{agreement} & : \left[ \begin{array}{ll} \text{gender} & : f \\ \text{number} & : s \end{array} \right] \\ \text{det} & : y \end{array} \right]$
--	-----	--

Чрез граф:



Пример на описание на правила:

$S \rightarrow NP VP$   $\langle NP \text{ person} \rangle = \langle VP \text{ person} \rangle$   $\langle NP \text{ number} \rangle = \langle VP \text{ number} \rangle$   
 $\langle S \text{ person} \rangle = \langle NP \text{ person} \rangle$   $\langle S \text{ number} \rangle = \langle NP \text{ number} \rangle$

В друг запис:

$$\begin{array}{c} S \\ \left[ \begin{array}{l} \text{cat} : s \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array} \rightarrow \begin{array}{c} NP \\ \left[ \begin{array}{l} \text{cat} : np \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array} \begin{array}{c} VP \\ \left[ \begin{array}{l} \text{cat} : vp \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array}$$

АС могат да описват синтактична информация.

УГ представят информацията чрез атрибути и техните стойности.

Програмата прави анализ по някакъв алгоритъм като на всяка стъпка, когато строи дървото на разбора към съответния възел (поддърво) се описва и съответната АС.

За думите стойностите на атрибутите са в речника (или се получават от морфологичен анализатор). За другите възли, те се получават чрез унификация на АС.

### III. Унификация на АС (от две или повече АС се получава нова)

Унификация на АС се дефинира по следния начин:

Две АС се унифицират, ако се унифицират стойностите на елементите им с еднакви атрибути.

Стойностите се унифицират по следните правила:

1. Променлива се унифицира с всяка стойност.
2. Константа се унифицира само с равна на нея константа или с променлива.
3. Атрибутна структура се унифицира с променлива или с атрибутна структура, като се приложат правилата рекурсивно.

Елементите, чиито атрибутни имена се срещат само в едната структура, се преписват в крайния резултат.

Примери

– когато две АС се унифицират:

$$\left[ \begin{array}{l} \mathbf{a} : a_1 \\ \mathbf{b} : \left[ \begin{array}{l} \mathbf{c} : c_1 \\ \mathbf{d} : d_1 \end{array} \right] \end{array} \right] \text{ унификация } \left[ \begin{array}{l} \mathbf{a} : X \\ \mathbf{b} : [\mathbf{d} : Y] \\ \mathbf{f} : X \end{array} \right] = \left[ \begin{array}{l} \mathbf{a} : a_1 \\ \mathbf{b} : \left[ \begin{array}{l} \mathbf{c} : c_1 \\ \mathbf{d} : d_1 \end{array} \right] \\ \mathbf{f} : a_1 \end{array} \right]$$

– когато две АС не се унифицират:

$$\left[ \begin{array}{l} \mathbf{a} : a_1 \\ \mathbf{b} : \left[ \begin{array}{l} \mathbf{c} : c_1 \\ \mathbf{d} : d_1 \end{array} \right] \end{array} \right] u \left[ \begin{array}{l} \mathbf{a} : X \\ \mathbf{b} : [\mathbf{d} : X] \\ \mathbf{f} : f_1 \end{array} \right]$$

#### IV. Правила и атрибутни структури

АС служат за дефиниране на ограничения за прилагане на правилата на граматиката.

За да опишем атрибутните структури за дадена граматика, трябва да определим *ограниченията, атрибутите и множествата от стойности* на всеки атрибут.

Например разглеждаме ограниченията:

- съгласуване по число (и лице) между подлог и сказуемо ("кучето лае", но не "кучето лаят"; "то лае", но не "ти лае");
- съгласуване по род и число между прилагателното и съществителното име ("бялото куче", но не "бялата куче" или "бялото кучета");
- категоризиране на глаголите в зависимост от това дали приемат допълнение или не, т.е. дали са преходни или непреходни ("гони" има допълнение, "лае" няма);
- само първата дума от именната група може да бъде членувана ("бялото куче", но не "бялото кучето").

За представяне на тези ограничения ни трябва следните атрибути:

лице	person	{ 1, 2, 3 }
число	number	{ s, p }
брой аргументи на глагола (преходност)	arg (или subcat)	{ 0, 1 } (или { 1, 2 })
род	gender	{ m, f, n }
членуване (определеност)	det	{ y, n } или { f, s, y, n }
категория	cat	{ n, v, adj, np, vp, s }

#### Представяне на думите

Речник - систематично и достъпно кодиране на разнообразна информация  
Всяка дума се представя в речника чрез АС, която кодира:

- синтактична информация
- семантична информация
- морфологични правила

Видове синтактична информация: част на речта, аргументи, присъщи синтактични характеристики, напр. род, число.

За да не се дублира информацията в речника се използват *макроси* за описание на обща информация за клас думи. Разширението на макроса се осъществява или при четене на думата, или при унификация. (Ако се извърши предварително се увеличава много речника.)

Може да се използват и стойности по подразбиране за атрибутите. (Напр., съществителните да се считат от 3 лице.)

Синтактичните характеристики могат да бъдат извлечени от морфологичен анализатор.

### Представяне на правилата

За да се приложи дадено правило, фразите от изречението, които се свързват със символите в правилото, трябва да отговарят на определени ограничения. Тези ограничения се задават чрез АС, прикрепени към символите в правилото.

Ограниченията са два типа:

- изискване даден атрибут да има конкретна стойност
- изискване атрибути в две или повече различни АС да имат равни (унифицируеми) стойности.

Освен това, след прилагане на правилото някои от атрибутите на символа вляво получават стойности от атрибутите в АС на символите вдясно.

Счита се, че описанието е частично – ако даден атрибут липсва, то описанието трябва да е съвместимо с коя да е стойност на този атрибут.

Пример за правила с АС:

$$\begin{array}{c} NP \\ \left[ \begin{array}{ll} \text{number} & : X \\ \text{det} & : Y \end{array} \right] \end{array} \rightarrow \begin{array}{c} Adj \\ \left[ \begin{array}{ll} \text{number} & : X \\ \text{det} & : Y \end{array} \right] \end{array} \begin{array}{c} N \\ \left[ \begin{array}{ll} \text{number} & : X \\ \text{det} & : n \end{array} \right] \end{array}$$

Използването на едно и също име за променлива тук показва, че стойността на атрибута number трябва да е една и съща. Същото се отнася и за стойността на "det".

В примера има ограничения от първия и от втория тип, освен това има предаване на стойност на атрибут в лявата част на правилото.

$$\begin{array}{c} NP \\ \left[ \begin{array}{ll} \text{number} & : X \\ \text{det} & : Y \end{array} \right] \end{array} \rightarrow \begin{array}{c} N \\ \left[ \begin{array}{ll} \text{number} & : X \\ \text{det} & : Y \end{array} \right] \end{array} \text{ Тук само се предават стойности.}$$

$$VP \rightarrow \begin{matrix} V \\ \left[ \begin{array}{l} \text{number} : X \\ \text{arg} : 0 \end{array} \right] \end{matrix}$$

Тук има ограничение от първия тип (конкретна стойност на атрибут) и предаване на стойности.

## V. Пример

Следният пример илюстрира използването на АС за описание на ограничения в граматиката.

Граматика:

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$NP \rightarrow N$

$NP \rightarrow Adj NP$

$N \rightarrow$  куче, кучето, кучетата, котка, котката, котки, котките

$V \rightarrow$  лае, лаят, гони, гонят

$Adj \rightarrow$  бялото, черната, черните

В речника:

съществителни

куче	кучето	кучетата
$\left[ \begin{array}{l} \text{cat} : n \\ \text{gender} : n \\ \text{number} : s \\ \text{det} : n \\ \text{person} : 3 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : n \\ \text{gender} : n \\ \text{number} : s \\ \text{det} : y \\ \text{person} : 3 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : n \\ \text{gender} : n \\ \text{number} : p \\ \text{det} : y \\ \text{person} : 3 \end{array} \right]$

котка	котката	котки	котките
$\left[ \begin{array}{l} \text{cat} : n \\ \text{gender} : f \\ \text{number} : s \\ \text{det} : n \\ \text{person} : 3 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : n \\ \text{gender} : f \\ \text{number} : s \\ \text{det} : y \\ \text{person} : 3 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : n \\ \text{gender} : f \\ \text{number} : p \\ \text{det} : n \\ \text{person} : 3 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : n \\ \text{gender} : f \\ \text{number} : p \\ \text{det} : y \\ \text{person} : 3 \end{array} \right]$

ГЛАГОЛИ

лае	лаят	гони	гонят
$\left[ \begin{array}{l} \text{cat} : v \\ \text{person} : 3 \\ \text{number} : s \\ \text{arg} : 0 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : v \\ \text{person} : 3 \\ \text{number} : p \\ \text{arg} : 0 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : v \\ \text{person} : 3 \\ \text{number} : s \\ \text{arg} : 1 \end{array} \right]$	$\left[ \begin{array}{l} \text{cat} : v \\ \text{person} : 3 \\ \text{number} : p \\ \text{arg} : 1 \end{array} \right]$

прилагателни

бялото		черната		черните	
cat	: adj	cat	: adj	cat	: adj
gender	: n	gender	: f	gender	: –
number	: s	number	: s	number	: p
det	: y	det	: y	det	: y

Правила:

$$1. \begin{array}{c} S \\ \left[ \begin{array}{l} \text{cat} : s \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array} \rightarrow \begin{array}{c} NP \\ \left[ \begin{array}{l} \text{cat} : np \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array} \begin{array}{c} VP \\ \left[ \begin{array}{l} \text{cat} : vp \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array}$$

$$2. \begin{array}{c} VP \\ \left[ \begin{array}{l} \text{cat} : s \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array} \rightarrow \begin{array}{c} V \\ \left[ \begin{array}{l} \text{cat} : v \\ \text{person} : X \\ \text{number} : Y \\ \text{arg} : 1 \end{array} \right] \end{array} \begin{array}{c} NP \\ [\text{cat} : np] \end{array}$$

$$3. \begin{array}{c} VP \\ \left[ \begin{array}{l} \text{cat} : s \\ \text{person} : X \\ \text{number} : Y \end{array} \right] \end{array} \rightarrow \begin{array}{c} V \\ \left[ \begin{array}{l} \text{cat} : v \\ \text{person} : X \\ \text{number} : Y \\ \text{arg} : 0 \end{array} \right] \end{array}$$

$$4. \begin{array}{c} NP \\ \left[ \begin{array}{l} \text{cat} : np \\ \text{person} : X \\ \text{number} : Y \\ \text{gender} : Z \\ \text{det} : D \end{array} \right] \end{array} \rightarrow \begin{array}{c} N \\ \left[ \begin{array}{l} \text{cat} : n \\ \text{person} : X \\ \text{number} : Y \\ \text{gender} : Z \\ \text{det} : D \end{array} \right] \end{array}$$

$$5. \begin{array}{c} NP \\ \left[ \begin{array}{l} \text{cat} : np \\ \text{person} : X \\ \text{number} : Y \\ \text{gender} : Z \\ \text{det} : D \end{array} \right] \end{array} \rightarrow \begin{array}{c} Adj \\ \left[ \begin{array}{l} \text{cat} : adj \\ \text{number} : Y \\ \text{gender} : Z \\ \text{det} : D \end{array} \right] \end{array} \begin{array}{c} NP \\ \left[ \begin{array}{l} \text{cat} : np \\ \text{person} : X \\ \text{number} : Y \\ \text{gender} : Z \\ \text{det} : n \end{array} \right] \end{array}$$

## Как работят правилата

Тъй като унификацията е комутативна операция, резултатът не зависи от реда на прилагането и правилата могат да се използват с всички типове анализатори.

Възходящ анализ на изречението: *Кучето гони черната котка.*

Възходящ анализ на изречението: *Бялото куче лае.*

Ако опитаме да анализираме "Бялото куче лаят.", няма да успеем, защото не можем да приложим правило 1 (единственото за S), тъй като атрибутът number има различни стойности за NP и VP.

## VI. Използване на УГ. Среди за представяне на УГ

Унификационните граматики се появяват в началото на 80-те години. УГ не са граматически теории, а среда за представяне на граматически теории. Някои граматически теории, базирани на тази идея, са Functional Unification Grammar (FUG, Kaplan 1985), Lexical Functional Grammar (LFG, Bresnan 1982), Generalized Phrase-Structure Grammar (GPSG, Gazdar, Klein, Pullum and Sag, 1985) и др. Разликата е в набора от атрибути, които използват.

Широкото използване на унификационните граматики е наложило разработването на специализирана среда за тяхното представяне - PATR. Тази среда е удобно средство за лингвистите при представяне на различни граматически теории чрез унификационни граматики.

PATR има език за представяне на правилата и ограниченията и интерпретатор на описаните в този език граматики.

Съществуват няколко подобрени версии - PATR II, DATR и др.

УГ са граматически формализми за представяне и на семантиката на изречение. За целта в АС се въвеждат семантични атрибути.

Пример: 
$$\left[ \begin{array}{ll} \text{word} & : \text{куче} \\ \text{cat} & : n \\ \text{sem} & : \# \text{куче1} \end{array} \right], \text{ където } \# \text{куче1} \text{ е указател към конкретен обект от даден клас в БД.}$$

Ако УГ се използват при превод от един език на друг, в АС на всяка дума е даден и преводът на втория език.

Пример: 
$$\left[ \begin{array}{ll} \text{lex} & : \text{маса} \\ \text{cat} & : \text{noun} \\ \text{agreement} & : \left[ \begin{array}{ll} \text{gender} & : f \\ \text{number} & : s \\ \text{det} & : n \end{array} \right] \\ \text{english} & : \left[ \begin{array}{ll} \text{lex} & : \text{table} \\ \text{cat} & : \text{noun} \\ \text{agreement} & : [\text{number} : s] \end{array} \right] \end{array} \right]$$