

РОСИЦА ДЕКОВА, СВЕТЛОЗАРА ЛЕСЕВА,
МАРИЯ ТОДОРОВА, ЕКАТЕРИНА ТЪРПОМАНОВА

**СТРУКТУРНИ ОСОБЕНОСТИ НА ФРАЗИ С ОПОРА
ПРИЛАГАТЕЛНО ИМЕ В БЪЛГАРСКИЯ ЕЗИК**

ROSITSA DEKOVA, SVETLOZARA LESEVA, MARIA TODOROVA,
EKATERINA TARPOMANOVA

**STRUCTURAL CHARACTERISTICS OF ADJECTIVE
PHRASES IN BULGARIAN**

(Abstract)

The paper summarises the major results obtained within the task for developing resources for demonstration taggers and lemmatisers carried out as part of the project *Integration of New Practices and Knowledge in the Education in Computational Linguistics*. The team's work involved the description of parts of Bulgarian grammar by means of formal syntactic rules with the further goal of implementing this formal description in a syntactic parser. As a result, we have devised a set of syntactic rules generalising on the structure of phrases headed by an adjective and their possible modifiers/complements – numerals, pronouns, adverbs and prepositional phrases. The rules were based on linguistic hypotheses attested by formal queries in the Bulgarian National Corpus.

Ke words: adjective phrases, Bulgarian, corpora studies, formal syntax

Статията представя част от резултатите, получени в рамките на задачата за разработване на демонстрационен софтуер за тагери и лематизатори по проект „Интегриране на нови практики и знания в обучението по компютърна лингвистика“¹. Работата на екипа беше свързана със създаване на формални правила, описващи част от българската граматика за целите на софтуерен инструмент за синтактичен анализ². В резултат беше съставен набор от синтактични правила, обобщаващи структурата на фраза с опора прилагателно име и различни разширения – числително име, местоимение, наречие и предложна фраза. Правилата се основават на езикови хипотези, засвидетелствани чрез формални заявки в Българския национален корпус (БНК)³, и са резултат от анализа на различни езикови примери. В тази статия са представени структурните особености на фрази от типа AP с разширения предложна група, числително име и местоимение.

1. Теоретична и метаезикова рамка

В нашето изследване разглеждаме термина „прилагателно“ в широкия смисъл на адектив и обхващаме не само прилагателните имена, традиционно определяни като „същински“, но и различни видове причастия, чиято синтактико-семантична функция съвпада с тази на прилагателните.

Предложеното формално описание на структурата на фраза с опора прилагателно име следва синтактичната теория на Пенчев (Пенчев 1993, Пенчев 1999), според която главни части в български могат да бъдат съществителни (N), прилагателни (A), глаголи (V), наречия (Adv), предлози (P/p) и съюзи (C/c), а за фраза се приема група думи, които могат да запълнят една синтактична позиция (Пенчев 1993: 19). Тестът за фраза, който е използван, е възможността за замяна с една местоименна дума (въпросително или лично местоимение).

Според Бъркалова (Бъркалова 2001) във фраза с опора прилагателно име (AP) прилагателните подчиняват съществителните чрез предложно свързване и прилагане. Комбинацията от две и повече прилагателни е възможна единствено в приименна позиция, затова наличието на комбинация AA, особено ако е съгласувана, изключва наличието на AP. Това води до извеждането на четири прости правила за AP (Бъркалова 2001: 31):

$$AP \rightarrow Adv A; AP \rightarrow ApN; AP \rightarrow AN; AP \rightarrow A (c) V = A (c) S,$$

които са обобщени с оглед на посоката на възможните разширения вляво и вдясно:

$$AP \rightarrow (Adv) A$$

Разгледаните и описани от нас комбинации във фраза с опора прилагателно име са: ApN (*достоеен за награда, гостуващ на семейството, най-силен сред момчетата*); ApPron (*уверен в някого, по-висок от някого, готов за това*); ApPronA (*по-висок от всеки друг*); AN (*дълъг километри*); ANumN (*висок два метра*).

2. Методология

За всяка една от възможните комбинации бяха разгледани и анализирани наличните срещания в Българския национален корпус (БНК) с цел откриване и дефиниране на правила за съгласуване във всеки тип подгрупа, както и определяне на границите на групата на прилагателното и установяване на евентуални маркери на фразовите граници (лява и дясна граница) за съответните типове. Съставени бяха хипотези, които съдържат терминалния символ (A), спрямо който се изследват началният и крайният за фразата символ и се описват категориите на фразовите компоненти. Верността и покритието на съставените правила се проверяват чрез езика за заявки в БНК или в други източници (предимно художествена литература). При наличие на конкретни ограничения и/или невъзможни комбинации, те също са дефинирани. По възможност са определени и маркери на

фразови граници, които могат да бъдат: дума, лема, част на речта (описвана както само по категория или като категория с конкретни граматични характеристики); списъци от затворени категории (лексикони) или препителни знаци.

Инструментът за създаване на правила позволява също да се определи коя дума е главната част, да се дефинира контекст (ляв и десен), както и да се зададе последователност на правилата. В статията представяме установените правила, илюстрирани с примери.

3. Синтактични правила, обобщаващи структурата на фраза с опора прилагателно име и различни разширения

3.1. Група на прилагателното с разширение предложна фраза (AP с разширение PP)

Прилагателните имена, които са опора във фраза с разширение предложна фраза, изпълняват предикативна функция в изречението. Фразата с опора прилагателно и разширение предложна фраза се описва от следните формални синтактични правила:

AP → A, PP (*висок до коленете, готов за бой*)

PP → P PP (*до над коленете*)

PP → P NP (*над коленете*)

Запетаята в първото правило означава, че словоредът на дадените конституенти не е фиксиран и следователно правилото описва както *висок до коленете*, така и *до коленете висок*.

Идентифицирани са прилагателни, които в определен контекст или значение изискват комплемент предложна фраза: *твърде непозната за едни кукли*; *твърде малка за такава уста*; *прекалено тежък за стомаха*; *напълно подходяща за горецината във влака*; *особено приятна за къпане*; *с вече приетите от Комисията предложения*; *(нещо) съвсем ново за мене*; *поддържат свой собствен защитен с парола профил*; *да бъдат общи за всички сектори*; *да създадем приятна за творчество атмосфера*; *почувствах се благодарен за думичката „ние“*; *притежаваше характерната за убийците самонадеяна суетност*; *едно дело, достойно за силите и таланта му*; *помощ, необходима за осигуряване на прехода*; *варианти, свързани с определянето на оперативните правила на МСЕ*; *лихвен процент, близък до разхода за заемане*; *при своята наета под наем колесница*; *увеличават своя пропорционален на БВП принос към бюджета на ЕС*; *осеяна с контролно-измервателни уреди и циферблати*; *старомодна чанта, пълна с ръждиви пръчки за голф*.

Установяването на инвентар от прилагателни и адективирани причастия, които изискват комплемент, както и на предлозите, които въвеждат комплемента на дадено прилагателно, е важно за определяне на това на

кое словосъчетание в даден контекст е конституент на дадена предложна фраза (вляво или вдясно от фразата, фраза, която е или не е в съседство).

В Българския национален корпус са засвидетелствани немалко на брой предложни фрази с повече от един (най-често с 2, но нерядко и с 3) предлога, поради което е включено и рекурсивно правило $PP \rightarrow P PP$.

Формулирани са лексикализирани правила за определени предлози в конкретни значения. Правилата са формулирани за даден първи предлог и възможните му комбинации с други предлози.

а) PP с първи предлог **до** със значение на темпорална, количествена или друга граница.

$PP \rightarrow \text{до} \{в, \text{над}, \text{около}, \text{по}, \text{под}, \text{след}\} NP$

$PP \rightarrow \text{до по} \{\text{над}, \text{около}, \text{под}, \text{след}\} NP$

Например: *константна **до в** рамките на 1 мг; високи **до над** коленете ботуши; усилен **до около** хиляда души; скрит **до след** полунощ;*

б) $PP \rightarrow$ първи предлог **на** със значение на количество или дистрибуция.

$PP \rightarrow \text{на} \{\text{до}, \text{между}, \text{над}, \text{около}, \text{по}, \text{под}\} NP$

$PP \rightarrow \text{на по} \{\text{между}, \text{над}, \text{около}, \text{под}\} NP$

Например: *отдалечен **на над** четиридесет левги; горящия **на около** миля от него пожар; равна **на около** 4000 м; разположени **на по** няколко сантиметра една от друга; разположени **на по над** 3200 метра надморска височина; пенсионирани **на по под** 50 г.; станали **на по между** 15–20 лв.; селото, разположено **на по около** 10 километра от двата града.*

в) PP с първи предлог **в** със значение на количество или дистрибуция.

$PP \rightarrow \text{в} \{\text{над}, \text{около}, \text{по}, \text{под}\} NP$

$PP \rightarrow \text{в по} \{\text{над}, \text{около}, \text{под}\} NP$

Например: *точен **в над** 95% от случаите.*

г) PP с първи предлог **от** със значение на количество или дистрибуция.

$PP \rightarrow \text{от} \{\text{над}, \text{около}, \text{по}\} NP$

$PP \rightarrow \text{от по} \{\text{над}, \text{около}\} NP$

Например: *образувана **от над** 4000 съставки; живеещите **от над** 200 г.; мъртви **от около** седмица; съставените **от по** трима души свързочни екипи; всеки от нас заобиколен **от по** двама надзиратели.*

д) PP с първи предлог **като** със значение на сравнение.

$PP \rightarrow \text{като} \{в, \text{за}, \text{на}, \text{над}, \text{по}, \text{пред}, \text{преди}, \text{през}, \text{при}\} NP$

$PP \rightarrow \text{като за} \{в, \text{пред}, \text{през}, \text{при}\} NP$

Например: *питомни **като в** селски двор; хубав **като за** лорд безгрижни и щастливи **като на** меден месец; същата **като преди** хиляди години; топло **като през** лятото; същият **като при** провал.*

Възможна е появата на втори предлог **за**, например: *широка **като за в***

хан, *скромно като за пред* Върховния септон, *жалки като за пред* крал, *облечен като за през* деня/зимата; *хумористични, като за през* Сирница.

- е) РР с първи предлог *за*, който въвежда комплемент.
 РР → за {преди, на, над, по} NP
 РР → за по {от, до, над, под} NP

Например: *подходящ за над* 14-годишна възраст; *добър за преди* тренировка.

- ж) РР с първи предлог *с* с количествено значение.
 РР → с {до, над, около, по} NP
 РР → с по {до, над, около} NP

Например: *подсладена с до* 5 на сто захари; *закъснялата с над* 18 месеца модернизация; *наклонена с около* единадесет градуса повече; *изпъстрена с по* няколко невъзпитани кучета; *повишени с по над* 100 лева.

Комбинациите от 3 предлога най-често включват дистрибутивния предлог *по* и означават съчетание от дистрибуция и друг признак (често количество), например: *издигнати на по над* осемстотин метра; *възлизащо на по над* 30 млн. долара, *увеличи бройката им с по около* триста души.

3.2. Група на прилагателното с разширение квантификатор (АР с разширение Num)

Група на прилагателното (АР) с разширение квантификатор има за опора прилагателно име, чиято семантика е свързана с размер или количество. В тези случаи АР е модификатор в именно словосъчетание NP, изразяващо мерни единици (за количество, размер и др.). Квантификаторното разширение в АР може да бъде фраза с опора съществително име (NP) или фраза с опора предлог (PP), които се описват от съответните формални синтактични правила.

3.2.1. Фраза с опора прилагателно име и именна фраза:

АР → NP, А – *висок двайсет и пет метра; двайсет и пет метра висок; километри дълъг; дълъг километри*

NP → NumP, N – *двайсет и пет метра; километри*

NP → N – *метра* във фразата *двайсет и пет метра; метри* във фразата *метри висок*

NumP → Num *двайсет*

NumP → Num Conj Num – *двайсет и пет*

NumP → NumP Conj NumP – *2 метра и 22 сантиметра*

3.2.2. Фраза с опора прилагателно име и разширение предложна фраза:

АР → PP, А – *с близо 2 метра по-къс*

PP → P PP – с по 3 метра по-широк

PP → P, NP – по-широк с три метра или с три метра по-широк

PP → PP PP – от 2 метра до 3 метра широк

В AP с разширение числително име често съставна част са уточняващи наречия (*само*), наречия за приблизителност (*близо*), наречия за долна граница (*поне*): *само до 2 метра висок; само около 2 метра широк; с близо 5 метра по-дълъг.*

3.3. Група на прилагателното с разширение местоимение (AP с разширение Pron)

В този случай разширението е от поне едно местоимение, което не е в кратка форма. При две или повече местоимения се наблюдават някои комбинаторни ограничения. Например показателните местоимения винаги бележат лявата граница на AP. Показателните, въпросителните, неопределителните, отрицателните и относителните местоимения не са съвместими помежду си, поради което са невъзможни комбинации като следните: *всеки този свой; всяка толкова своя; нечий такъв свой; какъв такъв свой; който такъв свой*. Не са възможни и комбинации между местоимения от един и същи вид (напр. комбинация от притежателно и възвратно-притежателно местоимение – *негов свой*).

При линейна комбинация, включваща пълна форма на възвратно притежателно местоимение, местоимението пояснява следходо съществително и не пояснява AP.

Фразата с опора прилагателно и разширение местоимение се описва от следното множество от формални синтактични правила:

1. Фраза с опора прилагателно име и разширение от едно или повече местоимения (показателни, неопределителни, въпросителни):

AP → Pron AP

AP → Pron* AP

такъв (някакъв/един) (много) умен

(някакъв/един/какъв) такъв (много) умен

При наличие на повече от едно местоимение се наблюдават описаните по-горе комбинаторни ограничения. Допуска се също и присъствие на наречие.

2. Фраза с опора прилагателно име и разширение от местоимение в дателен падеж:

AP → Pron_dat A

Обобщително, неопределително или отрицателно местоимение в мъжки род дателен падеж (*всякому, всекому, някому, никому*) може да бъде

разширение в AP, като може да предхожда или да следва прилагателното (*никому неизвестен, известен всекиму*), например: **Никому неизвестната** чиновничка Светла Тодорова вече ни струва стотици милиони, а скоро може да прескочи и милиард. Ако левият контекст е глагол, който присъединява непряко допълнение, комбинацията от дателно местоимение и следходно прилагателно може да не е AP, напр. *Не отстъпва никому пилотското кресло при кацане там.*

3. Фраза с опора прилагателно име и дясностояща, поясняваща го предложна група, която съдържа местоимение

Фразата включва местоименно NP със следните ограничения – NP да бъде пълна форма на личното местоимение във винителен падеж, пълна форма на възвратното местоимение, отрицателно, обобщително или показателно местоимение съществително.

AP → A PP

PP → P Pron – *уверен в себе си / него, разчитащ на себе си / мен*

4. Фраза с опора прилагателно в сравнителна степен

В този случай прилагателното в сравнителна степен се явява опора на по-голяма фраза – сравнителна конструкция. Обектът, с който се сравнява, се въвежда с предложна фраза, чиято опора е предлогът *от*.

AP → по- A PP – *по-висок от мен/ себе си/ всеки/ някой/ този*

PP → P Pron

P → от

5. Фраза с опора местоимение прилагателно

Задължително условие е местоимението да е в пълна форма. Когато въпросното AP е обособено, то се явява вдясно от поясняваното съществително и е в неопределена форма:

AP → Adv*, Pron

AP → Pron PP

Примери: *тази къща е наистина нейна
къщата, вече тяхна завинаги
къщата, негова след смъртта на чичо му*

Заклучение

В статията са представени резултатите от анализа на ексцерпиран езиков материал от Българския национален корпус и интернет, въз основа на който са изведени синтактични правила, обобщаващи структурата на словосъчетание с опора прилагателно име и различни разширения – числително име, местоимение и предложна фраза. Формалното представяне въз основа на правила може да бъде използвано за целите на автоматичния синтактичен анализ за български.

БЕЛЕЖКИ

¹ Проект „Интегриране на нови практики и знания в обучението по компютърна лингвистика“ (BG051PO001-3.3.06-0022 от 19.03.2012 г.) е съфинансиран от Европейския социален фонд на Европейския съюз по Оперативна програма „Развитие на човешките ресурси 2007–2013“ в рамките на схемата за безвъзмездна финансова помощ *Подкрепа за развитието на докторанти, постдокторанти, специализанти и млади учени*

² Natural Language Toolkit (<http://www.nltk.org/>)

³ Българският национален корпус е създаден в Секцията по компютърна лингвистика към ИБЕ при БАН и е достъпен на адрес: <http://search.dcl.bas.bg/>

ЛИТЕРАТУРА

Бъркалова 2001: *Бъркалова, П.* Комбинаторен извод на конституентните правила в съвременния български език. – В: Коева (съст.) *Съвременни лингвистични теории. Помагало по синтаксис.* Пловдивско университетско издателство.

Пенчев 1993: *Пенчев, Й.* Български синтаксис. Управление и свързване. Пловдивско университетско издателство.

Пенчев 1999: *Пенчев, Й.* Синтаксис. – В: *Бояджиев, Т., И. Куцаров, Й. Пенчев.* Съвременен български език. София, Изток–Запад, с. 498–655.

✉ *Гл. ас. д-р Росица Декова*

✉ *Гл. ас. д-р Светлозара Лесева*

✉ *Гл. ас. д-р Мария Тодорова*

✉ *Доц. д-р Екатерина Търпоманова*

Секция по компютърна лингвистика

Институт за български език „Проф. Л. Андрейчин“ при БАН

бул. „Шипченски проход“ 52, бл. 17, 1113 София, България

rosdek@gmail.com

zarka@dcl.bas.bg

maria@dcl.bas.bg

tochitsaaa@gmail.com

✉ *Assist. Prof. Rositsa Dekova, PhD*

✉ *Assist. Prof. Svetlozara Leseva, PhD*

✉ *Assist. Prof. Maria Todorova, PhD*

✉ *Assoc. Prof. Ekaterina Tarpomanova, PhD*

Department of Computational Linguistics

Institute for Bulgarian Language, Bulgarian Academy of Sciences

52 Shipchenski prohod blvd., bl. 17, 1113 Sofia, Bulgaria

rosdek@gmail.com

zarka@dcl.bas.bg

maria@dcl.bas.bg

tochitsaaa@gmail.com