

ФОРМАЛИЗИРАН МОРФОСИНТАКТИЧЕН РЕЧНИК НА ГЛАГОЛНИ ФРАЗЕОЛОГИЗМИ В БЪЛГАРСКИЯ ЕЗИК

Мария А. Тодорова

Резюме. Студията представя формализиран речник на морфосинтактичните характеристики на 1000 български фразеологизма с глаголна опора. Речникът е създаден за целите на автоматичната обработка на текстове и на компютърната лингвистика. Лингвистичната информация следва речниковия формализъм DELA, разработен от М. Грос. Формализираното описание комбинира категориална и граматическа информация със система от правила, представящи различни синтагматични характеристики.

Abstract. The paper presents a formalized dictionary of the morpho-syntactic characteristics of Bulgarian verb idioms aiming at their preprocessing in text. The dictionary comprises 1,000 lemmas of Bulgarian verb idioms. Their formalized description is covering categorical and grammatical information and a system of rules covering their syntagmatic characteristics. The linguistic information follows the DELA formalism, created by M. Gross.

Keywords: verb idioms, electronic dictionaries, Bulgarian language, natural language processing, language technologies

1. Увод

Качеството на лингвистичните методи за автоматично разпознаване на устойчиви словосъчетания зависи от прецизността на езиковите ресурси и модели, вложени в тях. Те включват различни по съдържание и обем речници и системи от правила, които описват лексикалните и граматичните характеристики на езиковите единици. Пълният списък на граматичните форми на фразеологизмите е необходим за компютърната обработка на даден език, но изписването на целите парадигми отнема много време и е свързано с дефиниране на конкретни комбинации в парадигматичните редове, които могат да се появят в контекста. Затова различните подходи за описание на фразеологични единици търсят закономерности, чрез които автоматично да възпроизведат някои от продуктивните им характеристики. А. Саг и др. (2002) изтъкват, че

ефективността на модела за описание на устойчивите словосъчетания зависи от групирането им според сходни характеристики и структура. Болдуин и др. (2003) посочват, че най-успешен за представяне на композираността в несвободните фрази е речниковият подход, съчетан с организирането им в краен брой класове според степента на семантична разложимост.

Настоящото изследване разглежда фразеологизми с опора глагол, наричани отгук нататък глаголни фразеологизми, независимо от семантичната им съотнесеност с определена част на речта. Изборът на този тип устойчиви фрази е мотивиран от няколко причини. Езиковите ресурси, представящи описание на устойчиви словосъчетания, както за български, така и за други езици, са насочени главно към именните устойчиви словосъчетания. Такива за български са например системата за разпознаване на съставни лексикални единици, предложена от Ив. Стоянова (2012); системата за класифициране на именувани същности, представена в Ив. Стоянова (2014); единиците, означени с фразеологичен маркер в БулНет¹ и аотирани като устойчиви словосъчетания в БулСемКор² (Тодорова и др. 2014) Другата причина да насочим интереса си към глаголните фразеологизми е широката им употреба и омонимията им със свободни фрази, която поражда трудности при автоматичната им обработка.

2. Глаголни фразеологизми

От гледна точка на компютърната лингвистика спецификите на глаголните фразеологични изрази произтичат от същността им на линейна последователност от графични думи. Необходимостта формалното описание да обхваща всички регулярни и нерегулярни форми на фразеологизмите предопределя изследването на техните парадигми и търсенето на последователни закономерности при синтагматичната им реализация. Разглеждаме парадигмата на даден фразеологизъм като съвкупност от всичките му реално употребими словоформи. В този смисъл фразеологичната словоформа е уникална последователност от микрословоформи³ към дадена фразеологична основна форма, която изразява уникално граматично значение.

3. Изисквания към формализираното представяне на глаголните фразеологизми

Изборът на формализирана рамка, която комбинира представянето на структурата на устойчивите словосъчетания и произтичащите от нея парадигматични и синтагматични специфики е предпоставка за създаването на

¹ <http://dcl.bas.bg/bulnet/>

² <http://dcl.bas.bg/semcor/bg/>

³ Терминът „микрословоформа“ е зает от Герджиков (1997). Тук, както и в предходни публикации, с него се означава всяка графична дума – конституент на даден фразеологизъм със свои морфологични особености.

качествен езиков ресурс, приложим в системите за автоматична обработка на езика. Основна цел при създаване на формализиран морфосинтактичен речник е детайлното и същевременно икономично описание на парадигмите на глаголните фразеологизми А. Савъри (2008) посочва следните критерии за оценка на качеството на различните формализми за представяне на флективни и морфосинтактични особености на устойчивите словосъчетания.

3.1. Критерии за пълнота

Описателните възможности, определящи за пълнотата на граматичното описание на даден формален подход към несвободните фрази, са:

- а) Икономичност на компонентното описание, което се постига чрез свързаност на морфологичните описания на самостоятелните думи и конституентите на несвободните фрази.
- б) Потенциал за морфологичен анализ и генериране на парадигмите на несвободни фрази, който обхваща: едновременно представяне на регулярни и дефективни парадигми; категоризиране на разделителите в рамките на несвободните фрази и описание на типови последователности във фразеологичната структура. Не на последно място е и капацитетът за представяне на флективните специфики на несвободните фрази като: екзоцентрични съгласувателни нерегулярности, деривационни варианти или възможни разширени или съкратени варианти.

3.2. Критерии за коректност

3.2.1. Нерегулярните съгласувателни характеристики са добре описани, ако са обхванати всички съществуващи форми, без да се въвеждат изкуствени лема.

3.2.2. Дефективните парадигми са добре описани, ако:

- а) не се наблюдава генериране или разпознаване на несъществуващи форми;
- б) лексикалните варианти на конституентите и вмъкването на допълнителни елементи могат да бъдат свързани с неутрална лема, която не ги съдържа;
- в) вариантите с незадължителен фразеологичен конституент са свързани с лема, която го съдържа.

3.2.3. Словоредните изменения са взети предвид, ако всички словоредни варианти се свързват с обща лема.

3.2.4. Деривационните и синтактични варианти трябва да бъдат свързани с минимална и неутрална изходна лема.

3.2.5. Големите флективни парадигми на устойчивите словосъчетания, обхващащи словоредните и съгласувателните им особености, са представени унифицирано.

3.2.6. Лемата на несвободната фраза не е абстрактна, ако е реално употребима в езика форма.

4. Формални подходи към морфосинтактичното описание на устойчиви словосъчетания

Специфичните парадигматични особености на фразеологизмите предопределят няколко начина, по които може да се осъществи формализираното им описание. Един от вариантите е създаване на речник, в който са изредени всички възможни парадигматични реализации на фразеологизмите, но такъв ресурс е твърде обемен и труден за реализиране. Според М. Грос (1996) лексикално-граматичните подходи са най-подходящата формална рамка за представяне на формообразуването при фразеологични единици. Те включват списък с основни форми и множество от правила, валидни за определени фразеологични групи. Тези правила се наричат локални граматика. Създаването на флективни речници и локални граматика може да се основава на различни математически формализми, пример за това са крайните автомати и базите от данни. Формализираното описание на фразеологизмите може да се изгради и според принципите на унификационните граматика.

4.1. Подходи, основани на релационни бази от данни

За морфологичното и синтактичното описание на устойчивите съчетания могат да се използват релационни бази от данни (Копстейк и др. 2002). Те позволяват интегриране на лингвистично описание от различни равнища в единен езиков ресурс и улесняват поддържането и достъпа до информацията. Релационните бази от данни съхраняват множество речникови единици под формата на таблици, записани като атрибути, и техните стойности в полетата. Те притежават редица преимущества, произтичащи от приложимостта им за представяне на различни формализми и възможността за съхранение на големи по обем данни. Релационните бази от данни не са речник, а инструмент за генериране на различни речници, който позволява детайлно описание, съчетаващо различни нива. Те могат да се използват в системи за изграждане на лингвистично знание⁴, които се прилагат както за разпознаване, така и за генериране, а някои от тях се разпространяват свободно като системи с отворен код. На релационни бази от данни е основан процесорът за устойчиви словосъчетания в баски – NAVIL (Алегрия и др. 2004). Той описва непоследователни, последователни, лексикализирани или декомпозируеми фразеологични форми.

4.2. Подходи, основани на унификационните граматика

Различните отношения на зависимост между фразеологичните микро-словоформи, като съгласуване и управление, определят необходимостта в описанието на фразеологизмите да бъдат включени едновременно морфологични и синтактични особености. Формализмът на унификационните грама-

⁴ Linguistic Knowledge Building, LKB

тики отговаря на това изискване. В него лингвистичните явления се описват чрез общи граматични правила, а специфичните за конкретни групи особености се кодират чрез отделни правила за лексикализация. Те се представят като отделен модул, който не оказва влияние върху езиковия модел, описващ регулярните единици. Практическо приложение на този подход може да се разгледа в проекта за синтактично и семантично описание на английски устойчиви словосъчетания LinGO (Cag и др. 2002; Копстейк и др. 2002; Вилавиченцио и др. 2004). В него синтактичните особености се представят чрез граматика на признакови структури, описани посредством опорно-фразов формализъм, основан на ограничения (Constraint Based Head Driven Phrase Structure Grammar). Друг, основан на унификационните граматика формализъм е FASTR – Fast Pater Term Recognizer (Жакмен 2001). Той представлява повърхнинен парсер за разпознаване, нормализиране и извличане на английски и френски компютърни термини. Програмата е специализирана за обработка на корпуси и за разпознаване на флективни синтактични и семантични варианти на съставни термини.

4.3. Подходи, основани на крайни автомати

При подходите, основани на крайни автомати, морфосинтактичните изменения на фразеологизмите се описват чрез множество от взаимно свързани и подредени според обхвата си морфологични и морфосинтактични правила. Крайните автомати могат да бъдат разпознаватели, които приемат или отхвърлят въведения входен низ, и преобразуватели, които променят входния низ според определени правила. Автоматите преобразуватели са устройства, които преработват входни данни в изходни. Ако се използват за обработка на езика, автоматите преобразуватели разпознават дадена езикова форма и я трансформират в друга езикова форма или в езикова форма и граматичната ѝ интерпретация на изхода. Формализмът на крайните автомати и преобразуватели се използва при няколко подхода за обработката на фрази. В тях се прилагат правила за преобразуване от една форма в друга. Такива са: френската програма Lexc (Картунен и др. 1992; Картунен 1993), в която крайните преобразуватели, представящи морфологичните изменения, са придружени от компилатор на регулярните изрази (Бийсли и Картунен 2003); програмата IDAREX (Брайд и др. 1996), разработена за описанието на немските глаголни несвободни фрази и техните вариации, и процесорът на несвободни фрази за турски (Multi-word processor) (Офлазер и др., 2004). В настоящото речниково описание е използван именно формализмът на крайните автомати.

5. Речниковият формат DELA

Разработеният в Лабораторията за автоматична и лингвистична обработка (Laboratoire d'Automatique Documentaire et Linguistique – LADL) към Френския национален център за научни изследвания (CNRS) формат DELA

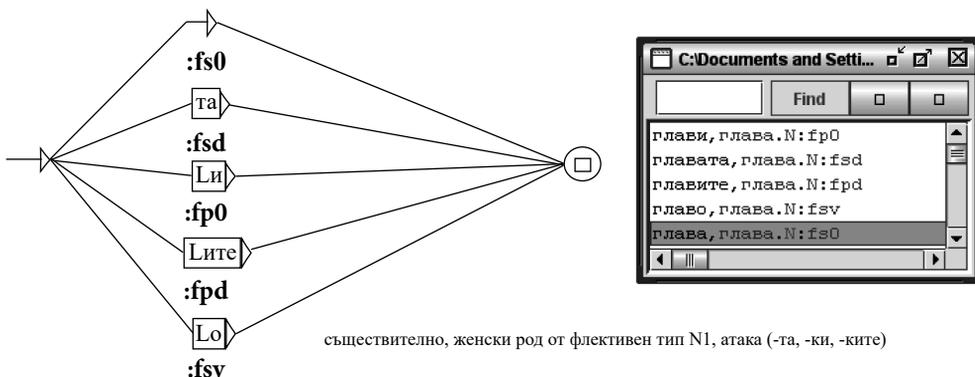
(Dictionnaires électroniques du LADL) (Куртоа и Зилберщайн 1990; Зилберщайн 1993а; 1993б) е насочен към широкообхватен морфологичен анализ на лексикални единици чрез автоматично съпоставяне на думи в текста с пълния списък от възможни граматични анотации. Автоматичното формообразуване на дадена лема се извършва с помощта на флективни граматика за различните флективни типове, в които се групират лемите. Връзката между лемите в речника и съответните им флективни граматика се осъществява чрез името на флективната граматика. Всяка флективна граматика се съотнася с точно определен флективен тип, чрез който се образуват формите в парадигмата и се приписва информация за граматичните категории и характеристики.

Разновидностите на DELA формата са различни речници, които могат да бъдат взаимосвързани. DELAS са електронни речници, които съдържат списък от лемите и техните флективни типове. В DELAS на всяка единица се приписват имената на флективни граматика, наричани още флективни кодове. Речниковата единица в DELAS има следния вид: битка, N4. С този запис на основната форма битка се приписва флективна граматика N4. Флективните граматика представляват крайни преобразуватели, които се представят чрез граф⁵ (фиг. 1). Всеки път в графа описва една или повече словоформи и представя последователност от преходи между състояния на преобразувателя. За удобство във формализма са въведени следните оператори, които съответстват на преходите между състоянията: *премахни последния символ, добави нов символ, премини един или n символа наляво или надясно*.

Приписването на флективни кодове на лемите в DELAS позволява автоматичното конструиране на DELAF речник. Той представлява списък от всички словоформи, лемата, към която принадлежат, и съответните им граматични характеристики. Автоматичното създаване на парадигмата на дадена лема се състои в образуването на всичките ѝ словоформи и характеризирането им със съответната граматична информация (фиг. 1).

DELA методологията се прилага при описанието на съставни и несъставни лексикални единици чрез DELAS и DELAC речници. Докато в DELAS се представят лемите на прости думи и съответстващите им флективни типове, в DELAC се представят лемите на лексикални единици, съставени от повече от една графична дума. Приписването на флективни кодове на всички единици в DELAC позволява автоматичното конструиране на DELACF речник, който съдържа всички словоформи и съответните им граматични характеристики.

⁵ Всеки автомат може да се представи като граф, например представяне на измененията на входна дума (напр. основна форма) чрез множество от преходи и изходи (до изходна словоформа).



Фиг. 1. Флективна граматика N4 (вляво), която образува словоформите на думата глава в DELAF (вдясно)

Синтактичните особености на простите и съставните думи се представят чрез локални граматик под формата на графи или таблици, като представянето е еквивалентно. М. Грос (1986) предлага модел за лексикално-граматично описание на несвободните фрази, при който те се разделят на класове и се характеризират като последователност от синтактични категории.

DELA формализмът осигурява възможност за представяне на възможна връзка между съставните лексикални единици и морфологичното описание на техните конституенти като отделни графични думи. Същевременно, чрез прилагането на флективни граматик словоформите се образуват автоматично. Други предимства са възможностите за представяне на флективната композираност на глаголните фразеологизми, отбелязването на опорен конституент – глагол, и на граматичния тип, към който принадлежи, описанието на структурни, словоредни и съгласувателни зависимости в рамките на израза.

DELA методологията е в основата на три от най-популярните, свободно достъпни системи за лингвистична обработка на езикови данни – текстове, речници и граматик: Intex (Зилберщайн 1993а; 1999а; 1999б), Unitex (Помие 2003), NooJ (Зилберщайн 2007), както и на специално разработената платформа за описание на устойчиви словосъчетания Multiflex (Савъри 2005; 2009). Комбинацията от различните модули в Unitex позволява различни решения за приписване на ортографични, синтактични и семантични варианти на несвободните фрази.

Така например в DELA версията за английски, разработена от А. Савъри (2000), съгласувателните особености се представят чрез унифициращи променливи – \$ и правила за изключения. В DELA версията на Кирякопулу (2002) за гръцки се описва генерирането на формите на устойчивите словосъчетания чрез прилагането на ограничителни филтри върху множеството от комбинации между формите на конституентите.

М. Зилберщайн (2002) предлага унифициран формализъм за описание на парадигмите на съставни и несъставни речникови единици в системата NooJ.

DELA операторите, приложени към лемите на простите думи, са допълнени с фразеологично ориентирани оператори, като *отиди в края на следващата дума* или *отиди до края на първия конституент*. По този начин всяка форма на дадена съставна единица се описва посредством поредица от преходи, извършвани върху суфиксите на конституентите. Формализмът на NooJ е удобен за описанието на конвенционални основни форми и на нерегулярни съставни единици, при чието формообразуване композираността не е съществена. Съставната фраза се разглежда като поредица от символи, включително интервалите, и не е необходимо отделните прости думи, които я съставят, да се описват поотделно. Флективните варианти могат да бъдат описани чрез съпоставяне на толкова поредици от оператори, колкото са необходими за определена форма. Недостатък на този подход е, че се пренебрегва композираността и формите на фразеологичните конституенти трябва да бъдат представяни многократно. Например описанието на устойчиво словосъчетание, съдържащо конституент *битка*, не може да разчита на парадигматичното описание на *битка* като отделна дума, а трябва да бъде допълнително разписван във всички съставни единици, в които е включен. Това създава трудности при описанието на „флективно богати езици“ като славянските.

Формализмът за описание на парадигмата на съставни думи Multiflex (Савъри 2009), имплементиран в Uniflex, използва унифициращ механизъм, чрез който се описват съгласувателните зависимости. Различните фразеологични конституенти се означават като променливи с различни стойности в локални граматика (Зилбершайн 1994; Грос 1996). Формализмът позволява коректното представяне на големи парадигми, на словоредни особености и на перифразни изменения.

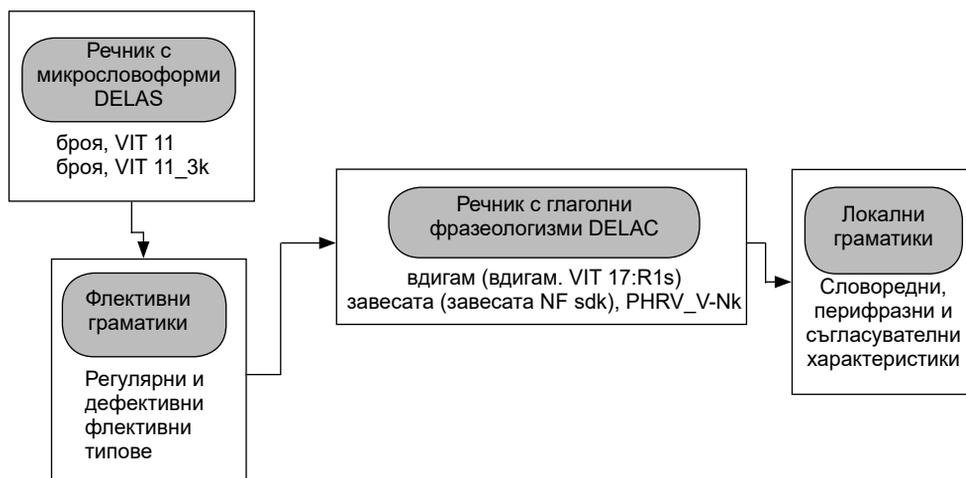
Чрез приписването на унифициращи променливи \$ на всяка от категориите (например \$v и \$n) микрословоформите могат да приемат всяка стойност от съответната граматична или лексикално-граматична категория – женски, среден или мъжки род; множествено или единствено число; бройна форма. Стойностите на категориите, които характеризират лемата, както и тези, с които се описва конкретната комбинация от микрословоформи, се преписват на изхода като характеристики на целия фразеологизъм. Този формализъм дава възможност за едновременно описание на словоредни, съгласувателни и флективни специфики и затова е особено удобен за флективно богати езици. Операторът за наследяване на стойности позволява прилагането на един и същи граф върху леми с еднакви структурни редове, но с различна флексия. По този начин се постига групиране на еднотипни фразеологични парадигми. Формализмът позволява изразяването на различни комбинации от морфологични стойности, включително некомпозирани и нерегулярни случаи. Изтриването, вмъкването и удвояването, както и други промени в конституентите се изразяват посредством дефинирания ред на променливите \$1, \$2, които реферират към отделните конституенти на лемата. В асоциираните DELACF речници всички флективни варианти са приписани на лемата.

6. Формализиран морфосинтактичен речник на български глаголни фразеологизми

Заради достъпността и възможностите за обхватност при описание на различни езикови нива в общ потребителски интерфейс за изработването на речник и граматика на български глаголни фразеологизми беше избрана системата Unitex⁶. Тя позволява комбинация от различни решения и модулни комбинации, подходящи за представяне на морфосинтактични специфики на устойчиви словосъчетания.

За да се постигне обхватно описание на регулярните и нерегулярни характеристики на избраното множество български глаголни фразеологизми, беше построено унифицирано лексикализирано описание. То обхваща речник на несъставните думи (микрословоформите), речник на фразеологичните единици, флективни граматика и локални граматика, представени като графи. Тяхното комбинирание, илюстрирано на фигура 2, води до автоматично образуване на възможните форми от парадигмата на всеки фразеологизъм, описан в речника на глаголните фразеологизми.

Всеки конституент в глаголния фразеологизъм се идентифицира морфологично в общ речник на несъставните думи. За да се генерират една или повече граматични форми на дадена фразеологична микрословоформа, е необходимо да знаем: нейната лема и флективния ѝ тип, означен с флективен код (т. 6.2.). Групирането във флективни типове се представя чрез флективни граматика (фиг. 3 и 4), които съответстват на краен автомат или преобразувател.



Фиг. 2. Структура на модулите на морфосинтактичното описание

⁶ <http://www-igm.univ-mlv.fr/~unitex/index.php?page=0>

Глаголните фразеологизми се описват чрез DELAC речник в Multiflex (т. 6.2.). Словоредните им особености се представят чрез правила за линейно предхождане на конституентите. В DELAC описанието те представляват локални граматика, които съответстват на краен автомат или преобразувател. Съгласувателните зависимости и перифразните трансформации, свързани с фразеологизирана позиция, вариациите на незадължителните компоненти в рамките на семантичен клас се групират в синтактични правила, които се прилагат допълнително при обработката на текст.

6.1. Съдържание на морфосинтактичния речник

За целите на настоящото изследване е създадена колекция от фразеологични данни (Тодорова 2009). Тя е извлечена от различни речници на българските фразеологизми, обработени по методологията за разработване на компютърния речник Phraseo-Lex на глаголни идиомати в немски (Фишер и Кайл 1996). Колекцията служи едновременно като източник за описание на фразеологични структури и за извличане на нови речникови единици. От нея е редуциран списък с глаголни фразеологизми, чиято опора се среща с голяма честота в източниците. Общият брой на глаголните фразеологизми в речника е 1000, броят на флективните типове в граматиката – 50, броят на типове перифрази – 5. В речника са представени 30 структурни типа.

6.2. Флективно описание на микрословоформите

Форматът на речника изисква представянето според общата Unitex структура, използвана в DELA – лексикална единица, формализирано описание. На всяка изменяема фразеологична микрословоформа се приписва граматична анотация. Граматичната анотация представлява лема и абстрактно описание, което съдържа флективен код. Флективният код реферира към граматиката, която описва съответните парадигматични редове. Граматиките, описващи регулярните флективни типове, следват описанието, представено в типове морфологични таблици на Б. Кръстев (1990). С условно възприетия индекс „k“⁷ се означава фразеологизираността на дадена микрословоформа в рамките на фразеологизма. С разделител двоеточие „:“ се изреждат граматичните характеристики на изменяемите граматично многозначни фразеологични конституенти⁸. Описанието им включва принадлежност към

⁷ Някои от означенията в студията се опростяват при имплементацията на описанието в Unitex.

⁸ Използваните означения за граматичните категории на лексемите и техните стойности са: N – съществително; A – прилагателно; NUM – числително; V – глагол; RD – сегашно деятелно причастие; DD – минало несвършено деятелно причастие; X – минало свършено деятелно причастие; P – страдателно причастие; Z – деепричастие; PRON – лично местоимение в именителен падеж; PRONme – лично местоимение във

граматична категория, граматичен тип, флективен тип и словоредни характеристики. Главните букви означават категория и категориално граматично значение, а малките – изменяеми граматични характеристики. Омонимните форми се описват едновременно. Представените регулярни флективни типове описват единиците, включени в речника, и не отразяват изчерпателно флективните типове в българския език. Групирането на местоименията в локални граматика е съобразено със структурните типове в речника (например фигура 8) и затова делението на регулярни и нерегулярни типове при тях е нерелевантно.

Таблица 1. *Разпределение на флективните типове в речника на словоформите DELAS*

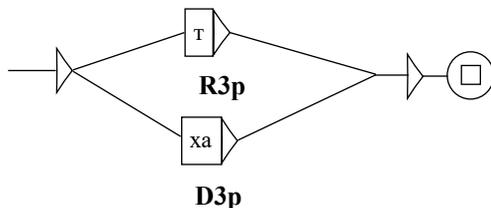
	Глаголи	Съществителни	Прилагателни	Местоимения
Общ брой флективни типове	28	21	8	15
Регулярни флективни типове	28	9	4	–
Фразеологизирани флективни типове	14	12	4	–

Незадължителни фразеологични елементи са думи от контекста или модификатори, които се вмъкват в минималната фразеологична конструкция. Задължителните позиции, отваряни от конструкциите, които не са запълнени с конкретна лексикална единица или множество от лексикални единици, се вписват с *някой* (N_1), *някого* (N_2), *нещо* (N_3).

Фразеологичните микрословоформи, чиято парадигма се различава от тази на регулярните думи, са включени с допълнителна дефективна флективна граматика в общия речник на несъставни единици DELAS. Флективният тип на неизменяемите микрословоформи съдържа само един член на парадиг-

винителен падеж; PRON_{mi} – лично местоимение в дателен падеж; PRON_{si} – възвратно лично местоимение; Q – въпросително местоимение; DEM – показателно местоимение; POS – притежателно местоимение; POS_{si} – възвратно притежателно местоимение; REL – относително местоимение; NON – неопределително местоимение; NEG – отрицателно местоимение; COL – обобщително местоимение.

Означенията на категориалните признаци са: а) при съществителни и прилагателни: M – мъжки род; F – женски род; N – среден род; б) при местоименията и причастията: m – мъжки род; f – женски род; n – среден род; в) при съществителни, прилагателни и местоимения: d – определена форма за женски и среден род; l – кратка определена форма за мъжки род; h – пълна определена форма за мъжки род; 0 – неопределеност; г) при местоименията: NOM – именителен падеж; ACC – винителен падеж; DAT – дателен падеж; w – пълна форма; q – кратка форма; д) при числителните: ORD – бройно; CAR – редно; е) при глаголите: T – транзитивност; I – интранзитивност; J – несвършен вид; P – свършен вид; R – сегашно време; D – минало несвършено време; E – минало свършено време; IMP – императивна форма; ж) други: s – единствено число; p – множествено число; 1, 2, 3 – лице.

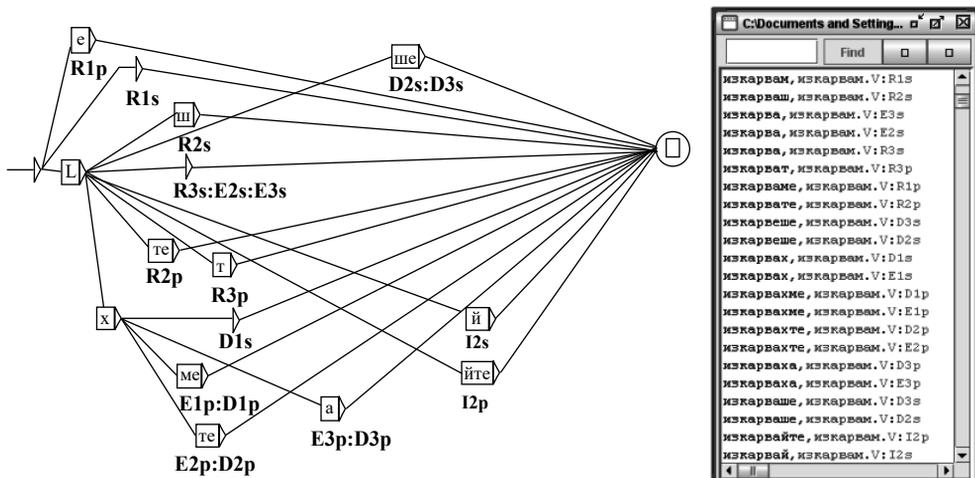


Фиг. 3. Флективна граматика, представяща дефективен флективен тип VIT11_3p_k

мата, който съвпада с основната форма. Ако парадигмата на регулярен глагол и фразеологичен глагол (опора във фразеологизъм) съвпадат, флективният код, използван в речниковото описание в DELAC е общ и реферира към една и съща флективна граматика. Така е например при глагола в *излизам от кожата си* 'ядосвам се', на който в DELAC е приписан флективният тип VIT15 с пълна реализация на всички словоформи – *излизам (излизам.VIT15:R1s) от кожата (кожата.NFsd_k) си, V_c-PREP_N_k-si. В броят ми се ребрата* опорната микрословоформа *броят* реферира към съответната редуцирана парадигма в речника VIT11_3p_k, в която глаголът се реализира само в 3 л. мн. ч. за личните форми (фиг. 3).

На фигура 4 е показана флективна граматика на транзитивни глаголи от несвършен вид, III спрежение, с окончание *-ам*. Автоматът на граматичния речник обхожда лемата и след това продължава по окончанията. Всяка дъга в графа означава различни операции, извършвани върху окончанието на ос-

изкарвам – преходен глагол от несвършен вид, от III спрежение, сег., мин св., мин несвр. вр.



Фиг. 4. Флективна граматика, представяща образуването на формите на микрословоформи с флективен тип VIT17a

новната форма: изтриване, добавяне, вмъкване на символи, за да се получи на изхода конкретна словоформа и съответстващото ѝ граматично описание. Например при прилагане на флективната граматика върху глагола *бързам*, с операцията *L – изтрий един символ от дясно наляво*, на изхода се получава формата *бърза* и ѝ се приписват граматичните характеристики *R3:E23s* – сегашно време, 3 лице; минало свършено време, 2-ро или 3-то лице, единствено число (посочени под състоянията).

6.3. Структурно и компонентно описание

Словоредните особености на фразеологизмите се представят чрез правила за линейно предхождане на конституентите. Както вече беше посочено, правилата се представят DELAC речник. В Multiflex структурните особености, словоредните вариации и типовите трансформации се описват чрез локални граматикати, а съгласувателните специфики се представят чрез задаване на стойности на променливите в локалните граматикати. Формализмът дава широки възможности за описание на съгласувателните особености, но е труден от потребителска гледна точка при представяне на фрази с много словоредни варианти, в рамките на които има съгласуване.

6.3.1. DELAC описание

Формализираното компонентно описание в DELAC включва абстрактната основна форма на всяка микрословоформа заедно със съответстващия ѝ флективен тип, както и код, описващ принадлежността на израза към даден структурен тип. Глаголните фразеологизми са представени според принципа за минималност и неутралност на абстрактната лема (Тодорова 2009). На всеки глаголен фразеологизъм се приписва код, който реферира към локална граматика. Фразеологичните локални граматикати едновременно описват различни словоредни подредби и съгласувателни характеристики (т. 6.5, 6.6, 6.7). В автомата към фразеологичните микрословоформи реферират номерирани променливи, означени със знака \$, например \$1, \$2, \$3 и т. н. Номерират се и интервалите, включени в структурата на фразеологизма.

6.4. Представяне на трансформационни еквиваленти

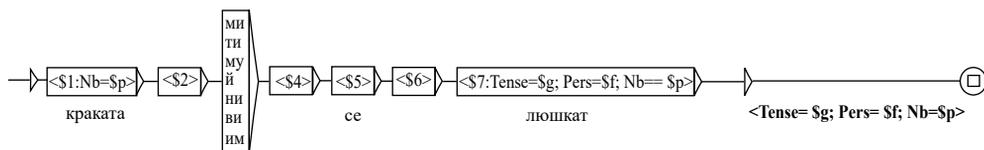
Всяка типова трансформация може да се представи чрез отделна граматика в краен автомат. Това позволява еднократното описание на различни словоредни размествания, както и на перифразни изменения като отрицателни конструкции, сравнителни конструкции, диатези, притежателни и обектни перифрази, които приемаме за еквиваленти на фразата в основна форма. Перифразните еквиваленти се приписват при структурните типове, при които се реализират. Те представляват синтактични правила и се представят чрез граматикати, в които са вградени други граматикати в Multiflex формализма. Прилагат се след аотирането на текста с DELAF речниците.

6.5. Съгласувателни характеристики

Съгласувателните зависимости се представят в Multiflex формализма чрез дефиниране на променливи $\$=...\$$ – *характеризира се с ... признаци*, и $\$=...\$$ – *приема стойностите на признаците на ...*, и чрез оператора $\$*...\$$, с който се задава ляв контекст, и $\$[...\$]$, с който се задава десен контекст.

Разглеждат се два типа съгласувателни зависимости – ендоцентрични и екзоцентрични.

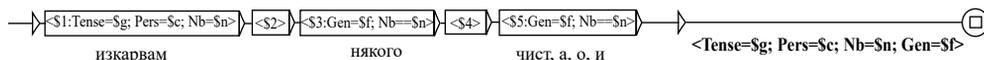
При ендоцентричното съгласуване опората управлява съгласувателните зависимости и определя допустимите синтактични отношения на целия фразеологичен израз. Определят се два типа ендоцентрични морфологични зависимости – контактни и дистантни. Означението $\$=$ показва, че променливата е представена във всички редове на парадигмата на съответния граматичен клас. В посочения на фигура 5 пример променливата $Nb=p$ показва съгласуването по число (множествено) на конституентите *люшкат* и *краката във фразеологизма краката ми се люшкат* ‘залитам’. Фразеологизмът приема стойностите за време на глагола, лице, число $\langle Tense=\$g; Pers=\$f; Nb=\$p \rangle$ в изходното морфологично описание на езика.



Фиг. 5. Формализирано представяне на ендоцентрично съгласуване

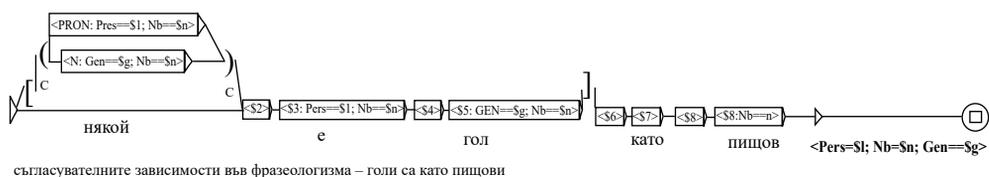
Екзоцентричното съгласуване се наблюдава при фразеологизми с малки изречения, които съдържат причастни или адвербиални микрословоформи в структурата си. Това определя описанието на съгласувателната им зависимост от думата, запълваща фразеологизираната субектна или обектна позиция, например *изкарвам някого чист/а/о*.

В посочения на фигура 6 пример променливите $Gen==\$f$ и $Nb==\$n$ показват задължителното съгласуване по род и число между прилагателното и думата във фразеологизираната позиция. В граматичното описание на фразеологизма се наследяват стойностите за лице, число и време на глагола и рода на прилагателното.



Фиг. 6. Формализирано представяне на екзоцентрично съгласуване

Едновременното екзо- и ендоцентрично съгласуване характеризира глаголни фразеологизми с малки изречения, чиито микрословоформи се съгласуват едновременно помежду си и с фразеологизираната позиция на подлога или допълнението. Например в гол/а/о/и като пищов съм ‘твърде беден съм’ и жив/а/о да го оплачеш ‘някой е много жалък’ прилагателното се съгласува с думата в подложната позиция и определя стойността на съществителното – *Той е гол като пищов; Те са голи като пищови*. В Multiflex формализма тези свойства се представят чрез променливите за операторите за ляв \$*...\$ и десен контекст \$[...\$], отбелязани на фигура 7 с кръгли и квадратни скоби.



Фиг. 7. Формализиране на частично независимото съгласуване

Както е в примера на фигура 7, при наличие на лично местоимение или съществително преди фразеологичния израз, контекстът обхваща прилагателното и спомагателния глагол и се задават еднакви стойности на съпадащите категории на трите конституента. Граматичните стойности, които се приписват на целия израз, са комбинация от стойностите на прилагателното и спомагателния глагол, независимо от наличието или липсата на контекст.

6.6. Перифразни трансформации

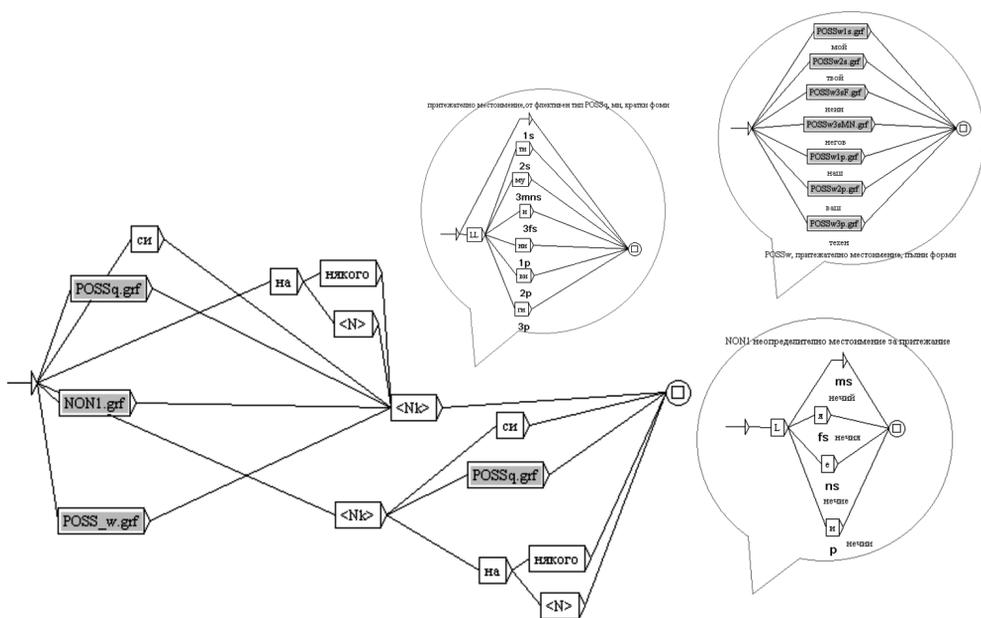
Едно от основните предимства на описанието чрез локалните граматик е възможността за относително свободно представяне на различни конструкции и вграждането на граматиката една в друга. Това позволява гъвкаво и икономично представяне. Специфични перифрази, които се реализират при определени структурни класове и типове, са:

а) **Притежателни перифрази**, които модифицират фразеологизирана позиция на прякото и непрякото допълнение на глаголната микрословоформа. Изразяват се чрез редуване на кратката форма на притежателното или възвратното притежателно местоимение с предложна фраза. Притежателни перифрази са и редуванията на пълната форма на притежателното местоимение или неопределителното местоимение за притежание с притежателна фраза. Приписват се при наличие на фразеологизирана притежателна позиция, запълнена от кратка форма на притежателното местоимение; от кратка форма на възвратното притежателно местоимение (si_N_k); от неопределително мес-

тоимение за притежание ($NON_{1_N_k}$) или от предложна фраза. Локалната граматика се вгражда в описанието на съответния структурен тип на глаголните фразеологизми. Притежателни перифрази допускат например структурните типове $V-(NON_{1_N_k})$; $V-PREP-(NON_{1_N_k})$ или $(NON_{1_N_k})-V3_t$.

Притежателна перифраза от тип 1 е тази, в която неопределително местоимение за притежание ($NON_{1_N_k}$) се преобразува в притежателно местоимение или предложна фраза. Например *вдигам нечие кръвно* ‘ядосвам някого’ → *вдигам неговото кръвно* → *вдигам на някого кръвното* → *вдигам му кръвното* → *вдигам кръвното на някого* → *вдигам кръвното му*.

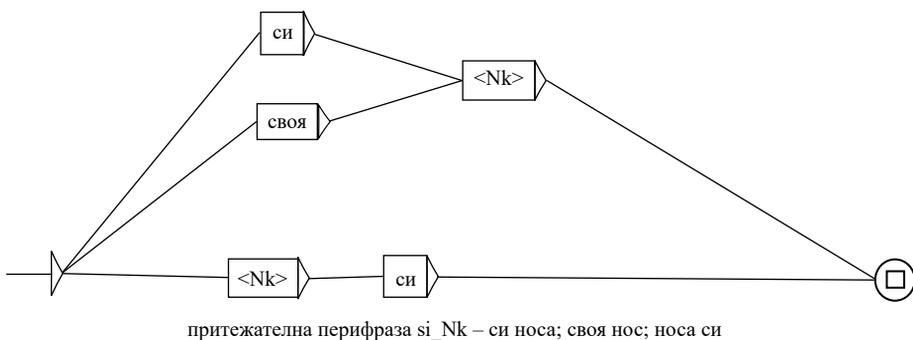
На фигура 8 са представени словоредните и конституентни варианти, свързани с перифразата. В локалната граматика ($NON_{1_N_k}$) са вградени локални граматика, които описват формите на неопределителното местоимение ($NON1$), кратките ($POSS_q$) и пълните ($POSS_w$) форми на притежателното местоимение. Съответно в локалната граматика $POSS_w$ са вложени локални граматика, които описват различните форми за лице на притежателното местоимение и съответните форми за род и число.



притежателна перифраза $NON1_N_k$ нечие кръвно; неговото кръвно;
на някого кръвното; му кръвното; кръвното на някого; кръвното му

Фиг. 8. Формализиране на притежателните перифрази на именна група ($NON_{1_N_k}$)

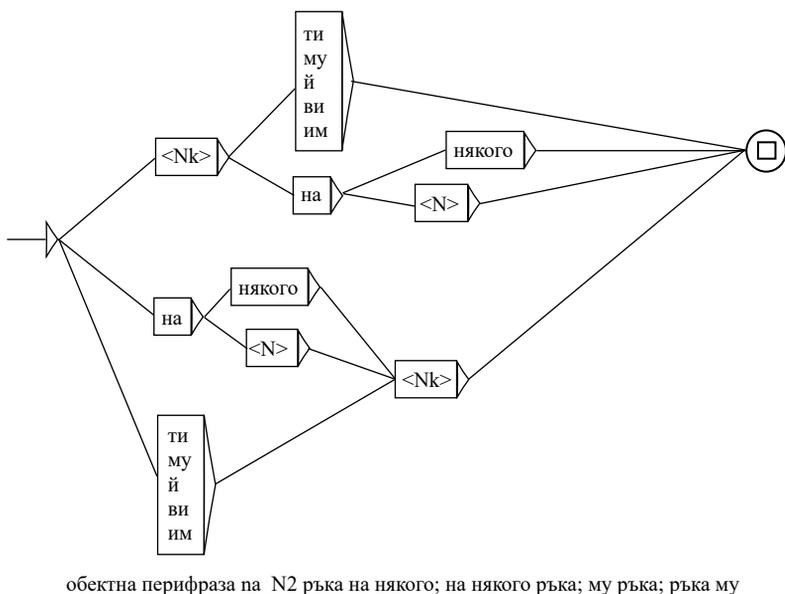
Възвратна притежателна перифраза от тип 2 (si_N_k), в която се редуват пълната и кратката форма на възвратното притежателно местоимение, е илюстрирана на фигура 9. Например *пъхам си носа* ‘нахалствам’ → *пъхам носа си* → *пъхам своя нос*.



Фиг. 9. Формализиране на притежателните перифрази на именна група (si_N_k)

Ограниченията при реализацията на конкретни комбинации като ограничението на едновременна реализация на едни и същи стойности на категорията лице при словоредна комбинация: глагол + кратка форма на притежателното местоимение **броя ми хапките* се представя експлицитно в локалната граматика.

б) **Обектни перифрази** са тези, които модифицират фразеологизирани позиции на непрякото допълнение на глаголните микрословоформи. Изразяват се чрез словоредни размествания, свързани с взаимната замяна на кратката форма на личното местоимение в дателен падеж с предложна фраза от типа *на+одушевено съществително*.



Фиг. 10. Формализиране на обектните перифрази на именна група

Приписват се на структурните типове $V-N_k-(na_N_2)$ или $V-PREP-N_k-(na_N_2)$. Например *подавам ръка на някого* ‘помагам’ → *подавам на някого ръка* → *подавам му ръка* → *ръка подавам на някого* → *ръка му подавам* (вж. фиг. 10).

в) **Сравнителни перифрази** са тези, които се изразяват чрез словоредни размествания, свързани с взаимната замяна на конструкции с предлога *като*, например *удрям с мокър парцал* ‘изненадвам неприятно’ → *като ударен с мокър парцал*.

г) **Условни перифрази** са тези, които се изразяват чрез словоредни размествания, свързани с взаимната замяна на конструкции от типа *животът ми е мил* ‘боя се’ → *мил ми е животът* → *ако ми е мил животът*; *живот и здраве да е* ‘пожелателна фраза’ → *ако е живот и здраве* → *нека е живот и здраве*.

д) **Отрицателни перифрази** са тези, които се изразяват чрез словоредни размествания, свързани със синонимната замяна на отрицателни конструкции, например *без да ми мигне окото* ‘без скрупули’ → *няма да ми мигне окото* → *окото ми не мигна*.

7. Разпределение на структурните типове

Броят, линейният ред и частта на речта на микрословоформите определят структурните класове; словоредните специфики и категориалните характеристики на микрословоформите определят структурните типове; флективните особености на неопорните микрословоформи определят структурните подтипове; флективният тип на опорния глагол определя фразеологични флективни типове.

Въз основа на концепцията за словоредни типове (Коева 2006) се разделят типови структури, които се реализират в различните подтипове: а) фиксиран словоред; б) граматично фиксиран словоред, представен чрез т.нар. перифразни еквиваленти; в) относително фиксиран словоред. Това деление се отразява и при флективното описание с условно приетите знаци: задължително празен интервал $_$; интервал, позволяващ вмъкване на клитики $-$; интервал, позволяващ вмъкване на конкретни класове или синтактични групи – наречия, прилагателни, адвербиални причастия и др. \sim .

7.1. Структурен клас $V+N_k$: фразеологизми с имплицитен подлог

Структурният клас обхваща глаголни фразеологизми, съставени от комбинацията на глагол и съществително.

7.1.1. Структурен тип $V-N_k$

Структурният тип включва транзитивни лични глаголи, в които позицията на прякото допълнение е фразеологично лексикализирана. Особеностите се определят от флективния тип на опората и от характеристиките на неопор-

ния конституент. Когато неопорните микрословоформи са фразеологизирани в една форма, те не се представят в отделни флективни граматика. Лексикализираната фразеологична микрословоформа се въвежда като отделна единица в речника на словоформите. В този тип фразеологизми глаголната опора се реализира с пълната си парадигма, представена чрез съответния флективен тип. Флективният тип обединява всички думи, които имат еднакво формообразуване, разбирано като флективни (и акцентни) редувания и съответните граматични характеристики на членовете на парадигмата. Глаголните фразеологизми от този тип подлежат на регулярните изменения на глаголите по лице, число и време.

Този структурен тип се характеризира с граматично фиксиран словоред, който допуска вмъкване на въпросителни клитики. Граматично фиксираният словоред се означава с „-“ в името на граматиката, която го описва – $V-N_k$.

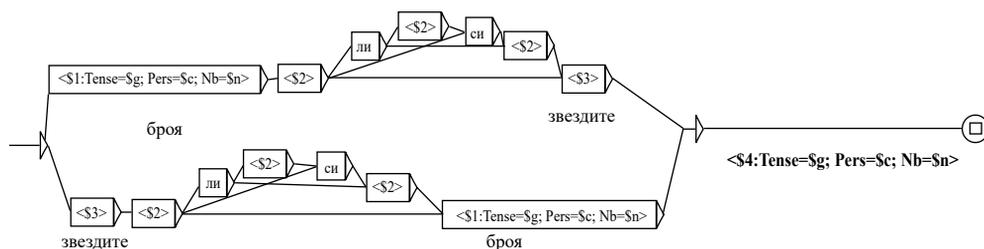
Групата се подразделя на структурни подтипове и съответно на флективни типове: $V-N0_k$ (*хвърлям котва* ‘установявам се’); $V-Nd_k$ (*броя звездите* ‘мечтая’). Структурните подтипове се определят от характеристиките на неглаголния конституент, а флективните типове от флективния тип на глагола. Към този структурен тип причисляваме и транзитивни лични глаголи, в които комбинацията с наречие е фразеологизирана.

В таблицата по-долу са представени разновидностите на структурния тип, който наброява сто двадесет и една основни форми в речника.

Таблица 2. Групиране на глаголните фразеологизми от структурен тип $V-N_k$

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
$V-NF_k$	$V-NFs0_k$	VIT17-NFs0 _k VIT8-NFs0 _k VIT18-NFs0 _k	27	<i>вдигам глава</i> <i>бера душа</i> <i>хвърлям котва</i>
	$V-NFsd_k$	VIT6-NFsd _k VIT17-NFsd _k	9	<i>бия камбаната</i> <i>вдигам махалата</i>
$V-NM_k$	$V-NMs0_k$ $V-NMsl_k$	VPT9-NMs0 _k VIT17-NMs0 _k VPT13-NMsl _k	73 14	<i>вдигам глас</i> <i>давам вид</i> <i>държа кокала</i>
	$V-NMpd_k$	VIT2-Npd _k	4	<i>вдигам гълъбите</i>
$V-NN_k$	$V-NNs0_k$	VIT17-NNs0 _k	6	<i>нямам сърце</i>
$V-N0/d_k$	$V-NMs0/d_k$	VIT6-NM_3 _k	4	<i>бия барабана</i>
$V-ADV_k$		VIT17-ADV _k VIT10-ADV _k	17	<i>гледам отвисоко</i> <i>изваждам наяве</i>

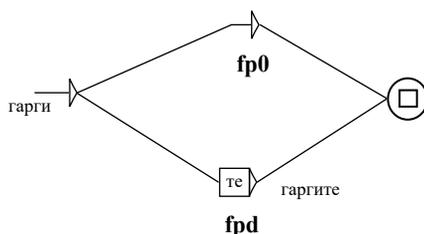
Словоредните изменения се описват с локална граматика, представяща граматично фиксиран словоред с два конституента – изменяем глагол и фразеологизирана микрословоформа. Представят се чрез локална граматика $PHRV_V-N_k$, илюстрирана на фигура 11.



$PHRV_V-N_k$ броя звездите; броя ли звездите; броя ли си звездите; звездите броя; звездите ли броя; звездите ли си броя

Фиг. 11. Локална граматика, представяща структурния тип $V-N_k$

В рамките на структурния тип се разглеждат и фразеологизми, при които неопорният конституент в позицията на фразеологизирано пряко допълнение е изменяем по един или повече граматични признаци. Изменяемите неопорни микрословоформи се представят в отделни флективни граматика, наречени дефективни (фиг. 12), тъй като не обхващат всички възможни форми на нефразеологичната лексема. Например в броя *гаргите* ‘разсейвам се’, $VIT11-NFp0d_k$ описанието на неопорния конституент *гаргите* отправя към дефективна граматика $NF1p_k$, в която съществителното *гарги* има само две форми, основна форма *гарги* и членувана форма *гаргите*.



фразеологизирано съществително, женски род от флективен тип $NF2pk$, гарги; гаргите

Фиг. 12. Локална граматика, представяща дефективен флективен тип

В бия *барабана* ‘вдигам шум’ $VIT6-NM_3_k$ описанието на неопорния конституент отправя към дефективна граматика NM_3_k , в която съществителното *барабана* има основна форма *барабан* и членувана форма *барабана*.

7.1.2. Структурен тип $V-A_k-N_k$

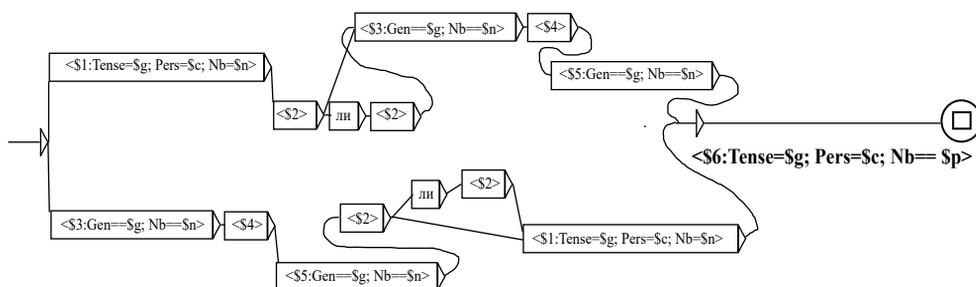
Структурният тип е разновидност на структурния клас $V+N_k$ и представлява комбинация от изменяем глагол, фразеологизирано прилагателно и фразеологизирано съществително, например *имам голяма уста* ‘говоря твър-

де много'. В речника тази група се състои от двадесет и шест основни форми и не изчерпва всички възможни случаи. На фигура 13 са представени словоредните и конституентни варианти.

Таблица 3. Групиране на глаголните фразеологизми от структурен тип $V-A_k-N_k$

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
$V-A_k-N_k$	$V-A_k-NFs0_k$	$VIT7-A_k-NFs0_k$	5	<i>имам дебела кожа</i>
	$V-A_k-NMs0_k$ $V-A_k-NMp0_k$	$VIT17-A_k-NMs0_k$ $VIT7-A_k-NMp0_k$	4	<i>давам заден ход</i> <i>имам гумени пръсти</i>
	$V-A_k-NNs0_k$	$VIT17-A_k-NNs0_k$	5	<i>имам лошо име</i>
$V-A_k-Ns0/l_k$	$V-A_k-NMs0/l_k$	$VIT17-A_4mlk$	4	<i>вдигам бял/белия байряк</i>
	$V-NUM_k-NFp0_k$	$VIT10-NUM_k-NFp0_k$	8	<i>права първи стъпки</i>

VC_V-Ak-Nk – имам дебела кожа; давам заден ход



Фиг. 13. Локална граматика, представяща структурния подтип $V-A_k-N_k$

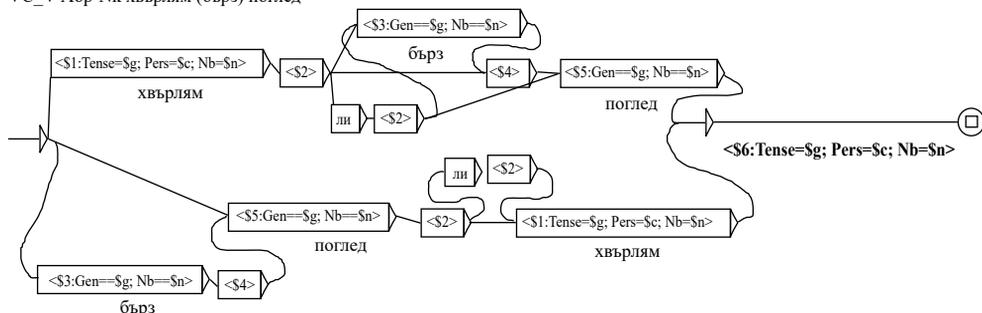
В рамките на структурния тип се разглеждат фразеологизми, при които неопорният конституент в позицията на фразеологизирано пряко допълнение е изменяем по един или повече граматични признаци и комбинации от изменяем глагол и фразеологизирани числително и съществително $V-Num_k-N_k$ (вж. таблица 3). Техните словоредни варианти също се представят с локалната граматика $V-A_k-N_k$ (фиг. 13).

7.1.3. Структурен подтип $V-(A_k)-N_k$

Структурният тип е разновидност на структурния клас $V+N_k$ и се състои от комбинация от изменяем глагол и фразеологизирана микрословоформа, между които може да се появи прилагателно, например хвърлям (бърз) поглед

‘преглеждам’, *бълвам (зелени) гущери* ‘злословя’. На този тип конструкции се приписва граматика, която описва ектоцентрично съгласуване между прилагателното и съществителното (фиг. 14). Разновидностите на структурния подтип се определят от флективните особености на микрословоформите, например *вдигам (голяма) пара* ‘гневя се излишно’ $V-(A)-N_{ok}$ и *бълвам (зелени) гущери* – $V-(A_{p0})-NM_{pok}$. В речника тази група се състои от четири основни форми и не изчерпва всички възможни случаи.

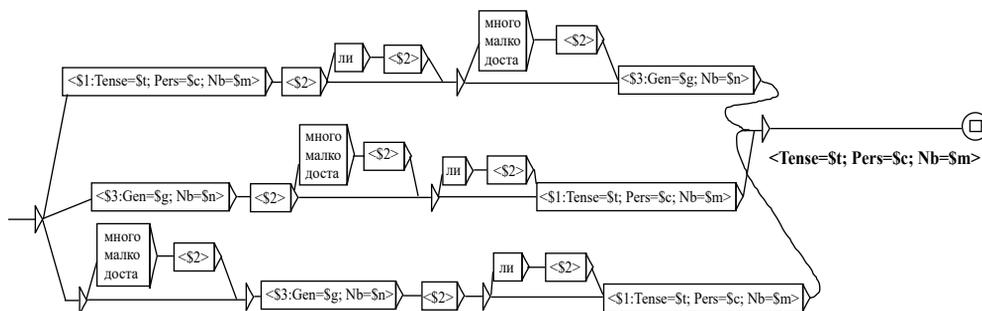
VC_V-Aop-Nk хвърлям (бърз) поглед



Фиг. 14. Локална граматика, представяща структурния подтип $V-(A_k)-N_k$

7.1.4. Структурен тип $V-(ADV_m)-N_k$

Структурният тип е разновидност на структурния клас $V+N_k$ и представлява комбинация от изменяем глагол и фразеологизирана микрословоформа, между които може да се появи наречие, например *виждам (много, малко, доста) зор* ‘затруднявам се’. На фигура 15 са представени словоредните и конституентни варианти. Тъй като наречieto варира в семантични класове, за всеки фразеологизъм от този тип се описва отделна граматика. В речника тази група се състои от четири основни форми и не изчерпва всички възможни случаи.



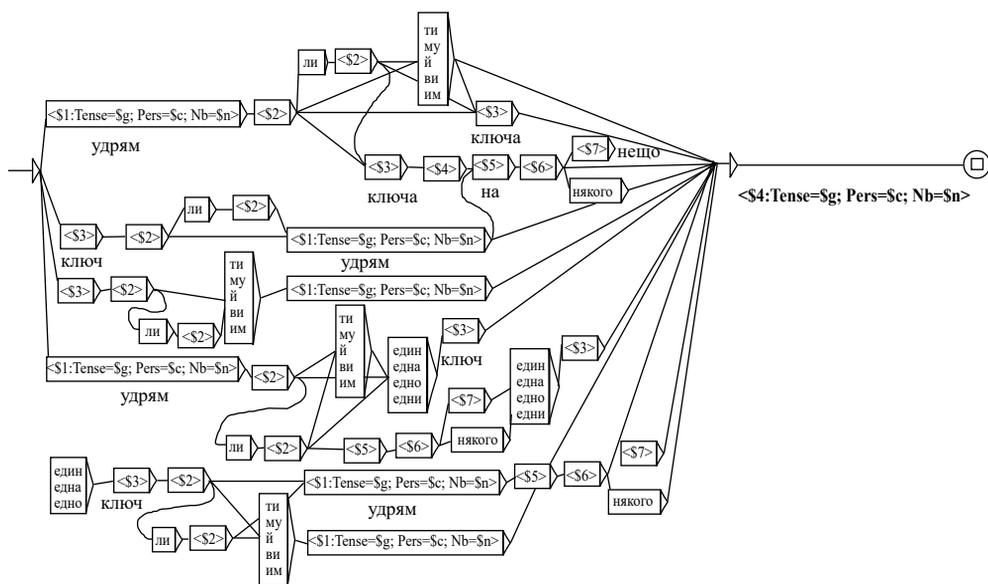
VC_V-ADV-Nk видях (много, малко) зор, зор малко ли видях, малко зор ли видях

Фиг. 15. Локална граматика, представяща структурния подтип $V-(ADV_m)-N_k$

7.1.6. Структурен тип V-(edin)-N_k (na_N₂)

Структурният тип е разновидност на структурния клас V+N_k и представлява комбинация от изменяем глагол и фразеологизирана микрословоформа, като *удрям ключа* ‘заключвам’, при които именният конституент може да бъде перифразиран с *един*, например *удрям един ключ*. На именния конституент, който има повече от една форма в парадигмата на фразеологизма, се приписва дефективен флективен тип. Глаголните фразеологизми от този тип се характеризират с граматично фиксиран словоред и обектна перифраза. В речника тази група се състои от шест основни форми и не изчерпва всички възможни случаи.

VC_V-(edin)-Nk_(na_N2) удрям ключ на нещо; удрям един ключ

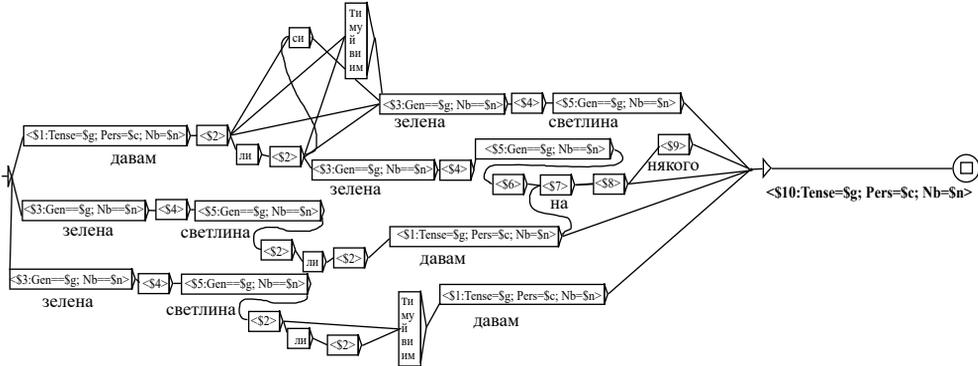


Фиг. 17. Локална граматика, описваща словоредата и перифразите на структурния тип V-(edin)-N_k-(na_N₂)

7.1.7. Структурен тип V-A-Nk_(na_N2)

Структурният тип е разновидност на структурния клас V+N_k и представлява комбинация от изменяем глагол, фразеологизирано прилагателно и фразеологизирано съществително, например *давам зелена светлина* ‘разрешавам’, и се характеризира с граматично фиксиран словоред и обектна перифраза. В речника тази група се състои от шест основни форми и не изчерпва всички възможни случаи.

VC_V-Ak-Nk-(na_N2) – давам зелена светлина на някого



Фиг. 18. Локална граматика, описваща словоред и перифразите на структурния тип V-A-Nk (na_N2)

7.1.8. Структурен тип V-(NON₁ N_k)

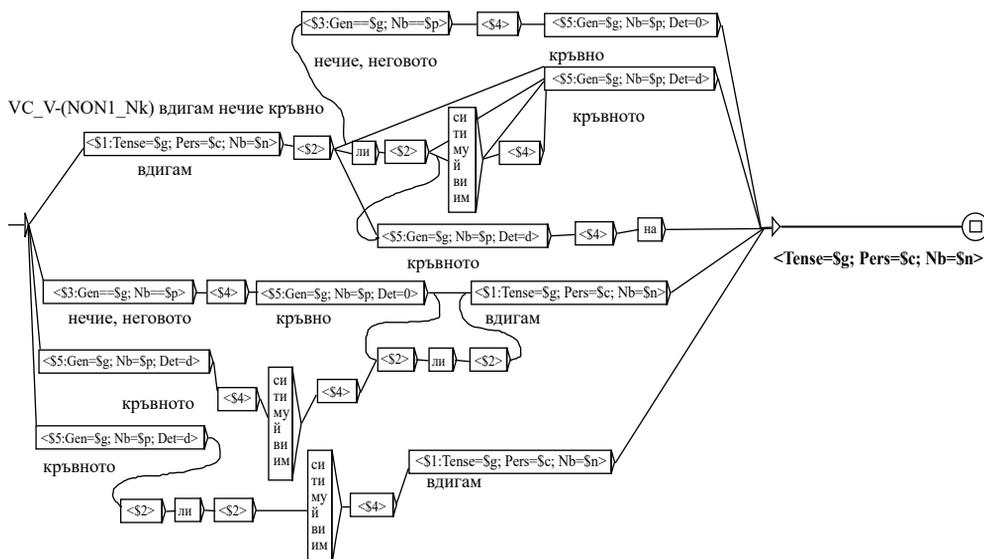
Структурният тип е разновидност на структурния клас V+N_k и се състои от транзитивни лични глаголи с фразеологизирано пряко допълнение и притежателна позиция. Глаголните фразеологизми от тази група, например *изтърся нечий джоб* ‘вземам парите’ се характеризират с граматично фиксиран словоред и притежателна перифраза от тип 1, която се приписва на съществителното (фиг. 9). В таблицата по-долу са представени разновидностите на структурния тип, който наброява четиридесет и шест основни форми в речника.

Таблица 5. Групиране на глаголните фразеологизми от структурен тип V-(NON₁ N_k)

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
V-(NON ₁ N _k)	V-(NON ₁ NF _s _k)	VPT10-(NON ₁ NF _{3s} _k)	3	<i>мътя нечия работа</i>
	V-(NON ₁ NM _s _k)	VIT10-(NON ₁ NM _{3l} _k)	2	<i>изтърся нечий джоб</i>
	V-(NON ₁ NMP _k)	V4IT-Npd _k -(NON ₁ NM _{7p} _k)	2	<i>ближа нечи крака</i>
	V-(NON ₁ NNS _k)	V17IT-(NON ₁ NNS _k)	5	<i>вдигам нечие кръвно</i>
	V-(NON ₁ NP _k)	V4IT-Npd _k -(NON ₁ N _p _k)	28	<i>ближа нечи сълзи</i>

Ограничението за реализацията на еднакви стойности на категорията лице при глагола и кратката форма на притежателното местоимение в 1 и 2 л. ?броя ми хапките; ?броиши ти хапките се представя в отделна граматика.

Ограничението отпада в трето лице – *брои му хапките* ‘наблюдавам някого колко яде’, и при употреба на възвратно притежателно местоимение – *броя си хапките*.



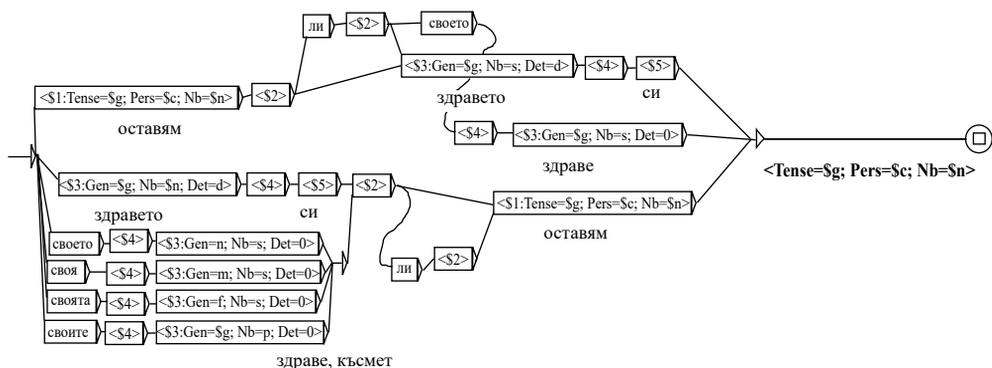
Фиг. 19. Локална граматика, описваща словоред и перифразите на структурния тип $V-(NON_1_Nk)$

7.1.9. Структурен тип $V-(N_k_POS_{si})$

Структурният тип обхваща транзитивни лични глаголи с фразеологизирано пряко допълнение и с фразеологизирана възвратна притежателна перифраза на именния конститuent. В речника тази група се състои от осемдесет и една основни форми и не изчерпва всички възможни случаи. Приписва се притежателна перифраза от тип 2 (вж. фиг. 20).

Таблица 6. Групиране на глаголните фразеологизми от структурен тип $V-(Nk_POS_{si})$

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
$V-(N_k_POS_{si})$	$V-(NFk_POS_{si})$	VIT17-NFsd _k _POS _{si}	34	<i>блъскам главата си</i>
	$V-(NMk_POS_{si})$	VIT17-NMsl _k _POS _{si}	39	<i>пъхам носа си</i>
	$V-(NNk_POS_{si})$	VPT18-NNsd _k _POS _{si}	2	<i>оставям здравето си</i>



VC_V-Nk_PRON si оставям (ли) здравето си; оставям своето здраве; здравето си (ли) оставям; своето здраве (ли) оставям

Фиг. 20. Локална граматика, описваща словоредата и перифразите на структурния тип $V-(N_k-POS_{si})$

7.1.10. Структурен тип $V-(N_k_POS_{si})-ADV$

Структурният тип обхваща транзитивни лични глаголи с фразеологизирано пряко допълнение, с фразеологизирана възвратно притежателна перифраза на именния конституент и наречие. В речника тази група се състои от 3 основни форми и не изчерпва всички възможни случаи, например *държа главата си високо* ‘запазвам достойнство’. Приписва се притежателна перифраза от тип 2, илюстрирана на фигура 9.

7.1.11. Структурен тип $V-(N_k_POS_{si})-A_k$

Структурният тип обхваща транзитивни лични глаголи с фразеологизирано пряко допълнение, с фразеологизирана възвратно притежателна перифраза на именния конституент и прилагателно, например *държа очите си отворени* ‘внимavam’. При тях се наблюдава ексцентрично съгласуване между съществителното и прилагателното, което се представя чрез стойностите на променливите, виж т. 6.5.

7.2. Структурен клас $V+PP$: фразеологизми с имплицитен подлог

Структурният клас обхваща лични глаголи с фразеологизирана предложна група. Структурният клас се подразделя на типове в зависимост от преходността на глагола. В речника тази група се състои от двеста тридесет и една основни форми и не изчерпва всички възможни случаи.

7.2.1. Структурен тип $V-PREP-N_k$

Структурният тип е разновидност на структурния клас $V+PP$ и се състои от интранзитивни лични глаголи с фразеологизирана предложна група.

Таблица 7. Групиране на глаголните фразеологизми от структурен тип V-PREP_N_k

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
V-PREP-N _k	V-PREP_NMs0	VII17-PREP_NMs0 _k	26	<i>умирам от студ</i>
	V-PREP_NMslk	VII17-PREP_NMsl _k	16	<i>влизам в строя</i>
	V-PREP_NFs0k	VII17-PREP_NFs0 _k	8	<i>падам в битка</i>
	V-PREP_Np0 _k	VII17-PREP_Np0 _k	15	<i>бия на очи</i>

7.2.2. Структурен тип V~N₂~PREP_N_k

Структурният тип е разновидност на структурния клас V+PP и се състои от транзитивни лични глаголи с фразеологизирана предложна група.

Таблица 8. Групиране на глаголните фразеологизми от структурен тип V~N₂~PREP_N_k

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
V~N ₂ ~PREP_N _k	V~N ₂ ~PREP_NMs0	VIT18~N ₂ ~PREP-NMs0 _k	8	<i>удрям (някого) в гръб</i>
	V~N ₂ ~PREP_NMslk	VIT18- N ₂ ~PREP_NMsl _k	14	<i>хвърлям (някого) в оркестъра</i>
	V~N ₂ ~PREP_NFs0 _k	VIT18- N ₂ ~PREP_NFs0 _k	17	<i>вземам (някого) на мушка</i>
	V~N ₂ ~PREP_NFsd _k	VIT18- N ₂ ~PREP_NFsd _k	7	<i>хващам (някого) за перушината</i>
	V~N ₂ ~PREP_NNs0 _k	VIT17- N ₂ ~PREP_NNs0 _k	16	<i>вземам (някого) за канарче</i>
	V~N ₂ ~PREP_A-N _k	VIT17- N ₂ ~PREP_An0 _k -NN0 _k)	19	<i>гледам (някого) с лошо око</i>
	V~N ₂ ~PREP_N _k CONJ_N _k	VIT17- N ₂ ~PREP_NM0 _k CONJ_NF0 _k)	12	<i>направлям (някого) на бъз и коприва</i>

7.2.3. Структурен тип V-PREP_(NON₁_N_k)

Структурният тип е разновидност на структурния клас V+PP и се състои от лични глаголи с фразеологизирана предложна група.

Таблица 9. Групиране на глаголните фразеологизми от структурен тип V-PREP_(NON₁ N_k)

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
V-PREP_(NON ₁ N _k)	V-PREP_(NON ₁ NFs0 _k)	VIII7-PREP_(NON ₁ NFs0 _k)	11	<i>влизам в нечия сметка</i>
	V-PREP_(NON ₁ NNs0 _k)	VIT17-PREP_(NON ₁ NNs0 _k)	17	<i>хвърлям в нечие лице</i>
	V-PREP_(NON ₁ Np0 _k)	VIII7-PREP_(NON ₁ Np0 _k)	18	<i>падам в нечи очи</i>
	V~N ₃ ~PREP_(NON ₁ Np0 _k)-N2	VIT17-N ₃ ~PREP_(NON ₁ Np0 _k)	6	<i>давам (нещо) в нечи ръце</i>

Подобно на структурен тип V-(NON₁ N_k) на подтиповете с фразеологизирано непряко допълнение се приписва притежателна перифраза от тип 1 на съществителното (фиг. 8).

7.2.4. Структурен тип (V_PART_{sc})-Prep-N_k

Структурният тип е разновидност на структурния клас V+PP и се състои от фразеологизирани възвратно лични глаголи. Флективният им тип съвпада с този на регулярния глагол и се съчетава с частицата *се*. Структурният подтип обхваща към момента тридесет и седем основни форми и се разделя според флективния тип и граматичните характеристики на микрословоформите.

Таблица 10. Групиране на глаголните фразеологизми от структурен тип (V_PART_{sc})-PREP-N

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
(V_PART _{sc})-Prep-N _k	(V_PART _{sc})-PREP-NF _k	VIT17Ref _k -se-PREP_NFsd _k	3	<i>бръквам се в кесията</i>
	(V_PART _{sc})-PREP-NM _k	VIT17Ref _k -se-PREP_NMsl _k	6	<i>бръквам се в джоба</i>
	(V_PART _{sc})-PREP-NN _k	VPT10Ref _k -se-PREP_NNsds _k	6	<i>бръкна се в портмонето</i>
	(V_PART _{sc})-PREP-Np _k	VIT6Ref _k -se-PREP_NFpd _k	4	<i>бия се в гърдите</i>
	(V_PART _{sc})-ADV	VIT13Ref _k -se-ADV	7	<i>държа се настрана</i>

В този тип обединяваме и фразеологизирани рефлексивни глаголи в комбинация с наречие като *държа се настрана* ‘избягвам’ и фразеологизирани рефлексивни глаголи в комбинация с предложна група.

Такива са структурен тип (V_PART_{sc})-Prep-(NON₁N_k) *хвърля се на нечия шия* ‘прегръщам’, както и фразеологизирани възвратни глаголи от структурен тип (V_PART_{si})-PREP_N_k (*набивам си в главата* ‘запомням’).

7.2.5. Структурен тип (V3t_PRON_{mc})-PREP_N_k и (V3t_PRON_{mi})-PREP_N_k

Структурният тип е разновидност на структурния клас V+PP и се състои от фразеологизирани безлични и третолични глаголи, при които парадигматичните изменения по лице се осъществяват от кратката форма на личното местоимение във винителен или дателен падеж. Обектната позиция може да се заеме от съществително, което задължително се удвоява. Описанието на глаголния конституент отправя към дефективна граматика V3t, в която глаголът има основна форма в 3 л. ед.ч. във всички времена. Структурният тип включва към момента пет основни форми и не изчерпва всички възможности. Такива са фразеологизирани безлични датива тантум (V3t_PRON_{mi})-PREP_N_k, *дреме ми на чепика* ‘не ме интересува’; *писва ми на шапката* ‘омръзва ми’; фразеологизирани безлични акузатива тантум (V3t_PRON_{mc})-PREP_N_k, *стяга ме под лъжичката* ‘притеснявам се’; *бодва ме в сърцето* ‘улучва слабото ми място’; *бодва ме под лъжичката* ‘внезапно изпитвам мъка’.

7.3. Структурен клас NP_k+V: фразеологизми с експлицитен подлог.

Класът се характеризира с фразеологизиран подлог, което определя дефективната парадигма на глаголната опора, съгласувана по число със съществителното в подложната позиция. Класът се разделя на следните типове.

7.3.1. Структурен тип (NON₁N_k)-VII3t

Структурният тип е разновидност на структурния клас NP_k+V и се състои от несвършени интранзитивни глаголи, фразеологизирани в трето лице. Структурният тип включва към момента четиринадесет основни форми и не изчерпва всички възможности. Характеризира се с притежателна перифраза от тип 1, например *нечие сърце изсъхва* ‘някой изпитва мъка’ → *изсъхва нечие сърце* → *неговото сърце изсъхва* → *изсъхва неговото сърце* → *изсъхва сърцето на някого* → *сърцето ми изсъхва*, и регулярно ограничение при реализацията на възвратното притежателно местоимение *?изсъхва си сърцето*.

Таблица 11. Групиране на глаголните фразеологизми от структурен тип (NON₁N_k)-VII3t

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
NON ₁ N _k -V3t	NON ₁ NNs0 _k -V3t	NON ₁ NNs0 _k -VII63t	1	<i>нечие сърце изсъхва</i>
	NON ₁ NMs0 _k -V3t	NON ₁ NMs0 _k -VII173t	13	<i>нечий дроб възвира</i>

7.3.2. Структурен тип (NON₁N_k)-CONJ_V3t

Структурният тип е разновидност на структурния клас NP_k+V и се състои от свършени интранзитивни глаголи, които се характеризират с притежателна перифраза от тип 1 и конституент съюза *да*, който може да бъде заменян с *ще*, например *нечие сърце да изсъхне* ‘пожелавам мъка’ → *да изсъхне нечие сърце* → *сърцето ми да изсъхне* → *сърцето на някого да изсъхне* → *нечие сърце ще изсъхне*.

Таблица 12. Групиране на глаголните фразеологизми от структурен тип (NON₁N_k)-CONJ_V3t

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
(NON ₁ N _k)-CONJ_V3t	(NON ₁ NNs0 _k)-CONJ_V3t	(NON ₁ NNs0 _k)-CONJ_VII63t	1	<i>нечие сърце да изсъхне</i>
	(NON ₁ NFs0 _k)-CONJ_V3t	(NON ₁ NFs0 _k)-CONJ_VIII173t	1	<i>приказка да излезе на някого</i>
	(NON ₁ NMs0 _k)-CONJ_V3t	NON ₁ NMs0 _k -CONJ_VII213t	11	<i>нечий дроб да възвие</i>

7.3.3. Структурен тип N_k-VII3t-(na_N₂)

Структурният тип е разновидност на структурния клас NP_k+V и се състои от несвършени интранзитивни глаголи, фразеологизирани в трето лице. Структурният тип включва към момента четири основни форми и не изчерпва всички възможности. Характеризира се с обектна перифраза.

Таблица 13. Групиране на глаголните фразеологизми от структурен тип N_k-(na_N₂)-VII3t

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
N _k -V3t-(na_N ₂)	NNs0 _k -V3t-(na_N ₂)	NNs0 _k -VII173t-(na_N ₂)	2	<i>име излиза на някого</i>
	NFs0 _k -V3t-(na_N ₂)	NFs0 _k -VII173t-(na_N ₂)	2	<i>приказка излиза на някого</i>

7.3.4. Структурен тип N_k -CONJ_V3t-(na_N₂)

Структурният тип е разновидност на структурния клас NP_k+V и се състои от свършени интранзитивни глаголи, които се характеризират с обектна перифраза и конституент съюза *да*, който може да бъде заменян с *ще*, например *име да излезе на някого* → *име да ми излезе* → *име ще излезе на някого*.

Таблица 14. Групиране на глаголните фразеологизми от структурен тип N_k -CONJ_V3t-(na_N₂)

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
N_k -CONJ_V3t-(na_N ₂)	NNs0 _k -CONJ_V3t-(na_N ₂)	NNs0 _k -VII173t-(na_N ₂)	1	<i>име да излезе на някого</i>
	NFs0 _k -CONJ_V3t-(na_N ₂)	NFs0 _k -VII173t-(na_N ₂)	1	<i>приказка да излезе на някого</i>

7.3.5. Структурен тип (NON₁N_k)-(PART_{sc-}V3t)

Структурният тип е разновидност на структурния клас NP_k+V и се състои от фразеологично рефлексивни интранзитивни глаголи. Именната микрословоформа се характеризира с притежателна перифраза от тип 1. Подразделя се според характеристиките на именния конституент и флективния тип на глагола. Включва към момента тридесет и три основни форми, описани в речника.

Таблица 15. Групиране на глаголните фразеологизми от структурен тип (NON₁N_k)-(PART_{sc-}V3t)

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
(NON ₁ N _k)-(PART _{sc-} V3t)	(NON ₁ Np _k)-(PART _{sc-} V3t)	(NON ₁ Np _k)-(PART _{sc-} VII113t)	18	<i>нечи кокали се броят</i>
	(NON ₁ NFs _k)-(PART _{sc-} V3t)	(NON ₁ NFs _k)-(PART _{sc-} VII12a3t)	10	<i>нечия глава се върти</i>
		(NON ₁ NMs _k)-(PART _{sc-} VII173t)	8	<i>нечий език се преплита</i>

7.3.6. Структурен тип N_k -(PART_{me-}V3t)

Структурният тип е разновидност на структурния клас NP_k+V и се състои от фразеологизирани третолични глаголи с фразеологизирана прякообектна позиция и фразеологизирано съществително. Структурният тип включва към

момента деветдесет и една основни форми. Характерна за типа е перифраза на прякообектната позиция с кратката форма на личното местоимение във винителен падеж, чрез която се осъществяват парадигматичните изменения по лице и число, например *студ ме побива*, *студ те побива* и т.н. Ако обектната позиция се изразява с име, то тя задължително се удвоява. Характеризира се с еноцентрично съгласуване между глагола и съществителното. Стойностите за число на променливите, означаващи съществителното и глагола, се унифицират. Словоредът е относително свободен и съществителното и глаголът може да се модифицират, съответно с наречие и с прилагателно. Характеризира се с обектна перифраза, например *студ ме побива* → *побива ме студ* → *студ побива някого*.

Таблица 16. Групиране на глаголните фразеологизми от структурен тип N_k -PART_{me}-V3t

Структурен тип	Структурен подтип	Флективен тип	Брой	Пример
N_k -PART _{me} -V3t	NMs0 _k -PART _{me} -V3t	NMs0 _k -PART _{me} -VII173t	8	<i>студ ме побива</i>
	NMsh _k -PART _{me} -V3t	NMsh _k -PART _{me} -VII63t	10	<i>бесът ме бие</i>
	NNs0 _k -PART _{me} -V3t	NNs0 _k -PART _{me} -VII123t	1	<i>куче ме влечи</i>
	NNsd _k -PART _{me} -V3t	NNsd _k -PART _{me} -VII103t	7	<i>шубето ме гони</i>
	NFs0 _k -PART _{me} -V3t	NFs0 _k -PART _{me} -VII123t	16	<i>мечка ме гази</i>
	NFsd _k -PART _{me} -V3t	NFsd _k -PART _{me} -VII103t	13	<i>душата ме боли</i>
	NFp0 _k -PART _{me} -V3t	NFp0 _k -PART _{me} -VII103t	10	<i>мравки ме лаят</i>
	NMpd _k -PART _{me} -V3t	NMpd _k -PART _{me} -VII93t	14	<i>дяволите ме вземат</i>
	A _k -N _k -PART _{me} -V3t	Afs0 _k -NFs0 _k -PART _{me} -VII173t	2	<i>студена пот ме избива</i>

7.3.7. Структурен тип N_k -POSS PART_{ne}-V3t

Структурният тип е разновидност на структурния клас NP_k+V и се състои от фразеологизирани третолични глаголи в отрицателна форма и фразеологизирано съществително. Описва се чрез отрицателни перифрази, в зависимост от вида на глагола, например *окото ми няма да мигне* → *няма да ми мигне окото* → *без да ми мигне окото* → *без окото ми да мигне*; *окото ми не мигва* → *не ми мигва окото* → *без да ми мигва окото* → *без окото ми да мигва*.

8. Обобщение

Електронните ресурси, описващи устойчиви словосъчетания, са от особена важност както за изграждането и разширяването на лексикалното покритие на електронните речници, така и за прецизирането и усъвършенстването на граматиките. Използването им цели подобряването на различни компютърни приложения като парсери, тагери, програми за автоматично съотнасяне на паралелни корпуси, системи за машинен превод.

В тази студия е представено приложението на морфосинтактичните характеристики на глаголни фразеологизми в конкретен формализъм за лексикализирано описание. Описанието предлага модел за типологизиране на глаголните фразеологизми с оглед на тяхната парадигма. Системата от модули в речниковия DELA формат, обединени в лексикализирана граматика, дава възможност за обхvatно и многостранно описание на глаголни фразеологизми. Тя е напълно адекватна на съвременните изисквания за качествени и функционални езикови ресурси. Създаденият формализиран речник има практическо приложение при автоматичното разпознаване и анализиране на продуктивни фразеологични типове. Той може да бъде преработван, конвертиран и прикачван към различни компютърни приложения. Форматът му позволява попълване с нови лексикални единици и с допълнителни граматични характеристики.

Практическата работа по лексикализирано описание в речник и автоматично извеждане на продуктивни типове на глаголните фразеологизми, както и създаването на обхvatен фразеологичен ресурс позволяват анализирането на разпределението на граматическата информация във фразеологичните форми. Организацията на езиковия материал потвърждава тезата за лексикализирането и граматикализирането на парадигматичните и синтагматични особености при фразеологизмите. Нерегулярността и аномалността не са свързани с образуването на фразеологичните единици, а се проявяват закономерно при тяхното функциониране. Дистрибуцията на различните перифразни преобразувания доказва, че те са характерни за конкретните структурни типове. Моделът е отправна точка за бъдещи изследвания, подлежи на досъвършенстване и е в процес на тестване върху обемен корпус за български език, както е Българският национален корпус⁹ (Коева и др. 2011; 2012) и изследване на резултатите.

Литература

Алегрия и др. 2004: Alegria, Iñaki, Olatz Ansa, Xabier Artola Nerea Ezeiza, Koldo Gojenola and Ruben Urizar. Representation and treatment of multiword expressions in Basque. – In: *Proceedings of the ACL 2004 Workshop on Multiword*

⁹ www.ibl.bas.bg/BGNK_bg.htm

Expressions: Integrating Processing, eds. Tanaka, Takaaki, A. Villavicencio, F. Bond, and A. Korhonen, Morristown, NJ: ACL, 2004. 48–55. <http://www.aclweb.org/anthology/W04-0407>.

- Бийсли и Картунен 2003:** Beesley, K., and L. Karttunen. *Finite State Morphology: Xerox tools and techniques*. Stanford: CSLI Publications, 2003.
- Болдуин и др. 2003:** Baldwin, T, C. Bannard, T. Tanaka, and D. Widdows. An empirical model of multiword expression decomposability. – In: *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment - Volume 18*. Association for Computational Linguistics Stroudsburg, PA, USA, 2003, pp. 89-96. <<http://anthology.aclweb.org/W/W03/W03-1812.pdf>>
- Брайд и др. 1996:** Breidt, E., F. Segond, and G. Valetto. Formal description of multiword lexemes with the finite-state formalism IDAREX. – In: *Proceedings of the 16th conference on Computational linguistics - Volume 2*. Association for Computational Linguistics Stroudsburg, PA, USA, 1996, pp. 1036-1040. <<http://www.aclweb.org/anthology/C96-2182?CFID=631180936&CFTOKEN=45776112>>
- Вилавиченцио и др. 2004:** Villavicencio, A., A. Copestake, B. Waldron, and F. Lambeau. Lexical encoding of MWEs. – In: *Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing*, eds. Tanaka, Takaaki, A. Villavicencio, F. Bond, and A. Korhonen, Morristown, NJ: ACL, 2004, pp. 80-87. <<http://www.aclweb.org/anthology/W04-0411>>
- Герджиков 1997:** Gerdžikov, G. Die Gesetze der gramatischen Paradigmata und die morphologische Typologie der Sprachen. – *Съпоставително езикознание*, 1997, 22(2), s. 30-41.
- Грос 1986:** Gross, M. Lexicon grammar. The representation of compound words. – In: *Proceedings of the 11th conference on Computational linguistics*. Stroudsburg, PA, USA, 1986, pp. 1-6. <<http://www.aclweb.org/anthology/C86-1001>>
- Грос 1996:** Gross, M. Lexicon-grammar. – In: *Concise Encyclopedia of Syntactic Theories*, eds. Brown, and J. Miller, Oxford: Pergamon Press, 1996, pp. 224-259.
- Жакмен 2001:** Jacquemin, Christian. *Spotting and Discovering Terms through Natural Language Processing*. Cambridge, Mass.: MIT Press, 2001.
- Зилберщайн 1993а:** Silberztein, M. *Dictionnaires électroniques et analyse automatique de textes: le système INTEX*. Paris: Masson, 1993.
- Зилберщайн 1993б:** Silberztein, M. Les groupes nominaux productifs et les noms composés lexicalisés. – *Linguisticae investigationes*, 1993, 17(2), pp. 405-425.
- Зилберщайн 1999а:** Зилберщайн 1999а: Silberztein, M. Text indexing with INTEX. – *Computers and the humanities*. 1999, 33(3), pp. 265-280. <<http://dblp2.uni-trier.de/db/journals/lre/lre33>>
- Зилберщайн 1999б:** Silberztein, M. INTEX: a finite state transducer toolbox. – *Theoretical computer science*. Elsevier Science 1999, 231(1), pp. 33-46.
- Зилберщайн 2002:** Silberztein, M. NOOJ: a cooperative object oriented architecture for NLP. – In: *INTEX pour la Linguistique et le traitement automatique des langues. Cahiers de la MSH Ledoux. 5th INTEX workshop, May 2002, Marseille, France*. Presses Universitaires de Franche-Comté, 2004. <https://www.researchgate.net/publication/266452836_NOOJ_A_Cooperative_Object_Oriented_Architecture_for_NLP>
- Зилберщайн 2007:** Silberzstein, M. *NOOJ manual*. <<http://nooj4nlp.net/>>, дата на достъп 01.09.2010 г.

- Картунен 1993:** Karttunen, L. *Finite-state lexicon compiler. ISTL-NLTT-1993-04-02* [Technical Report]. Palo Alto: Xerox Corporation Research Center, April 1993.
- Картунен и др. 1992:** Karttunen, L., R. M. Kaplan, and A. Zaenen. Two-level morphology with composition. – In: *Proceedings of the 14-th International Conference on Computational linguistics (COLING '92). Vol. 1.* Association for Computational Linguistics, 1992, pp. 141-148. <<http://www.aclweb.org/anthology/C92-1025>>
- Кирякопулу 2002:** Kyriacopoulou, T. Un système d'analyse de textes en grec moderne: représentation des noms composés. – In: *Actes du 5ème Colloque International de linguistique Grecque (Sorbone, 2001).* Paris : L'Harmattan, 2002, pp. 51-54.
- Коева 2006:** Коева, S. Inflection morphology of Bulgarian multiword expressions. – In: *Computer applications in Slavic studies – proceedings of Azbuki@net, International Conference and Workshop.* Sofia, 2006, pp. 201-216.
- Коева и др. 2011:** Коева, Св., Д. Благоева и Сия Колковска. Проектът Български национален корпус – резултати и перспективи. – *Български език*, 2011, 58(3), с. 34–53.
- Коева и др. 2012:** Коева, Sv., I. Stoyanova, Sv. Leseva, T. Dimitrova, R. Dekova, and E. Tarpomanova. The Bulgarian national corpus: theory and practice in corpus design. – *Journal of language modelling.* 2012, (1), pp. 65-110. <<http://jlm.ipipan.waw.pl/index.php/JLM/article/view/33>>
- Копстейк и др. 2002:** Copestake, A., et al. Multiword expressions: linguistic precision and reusability. – In: *Proceedings of the 3-rd International Conference on language resources and evaluation (LREC 2002),* 2002, pp. 1-7.
- Куртоа и М. Зилберщайн 1990:** Courtois, B., and M. Silberstein. Dictionnaires électroniques du français. – *Langue française*, 1990, 87(1), pp. 3-4. <http://www.persee.fr/doc/lfr_0023-8368_1990_num_87_1_6321>
- Кръстев 1990:** Кръстев, Б. Морфологията на българския език в 187 типови таблици. 2. изд. София: Наука и изкуство, 1990.
- Офлазер и др. 2004:** Oflazer, K., Ö. Çetonoglu, and B. Say. Integrating morphology with multi-word expression processing in Turkish. – In: *Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing,* eds. Tanaka, Takaaki, A. Villavicencio, F. Bond, and A. Korhonen, Morristown, NJ: ACL, 2004, pp. 64-71. <<http://www.aclweb.org/anthology/W04-0409>>
- Помие 2003:** Paumier, S. *UNITEX 2.1 user manual* [Electronic document]. Université Paris-Est Marne-la-Vallée, 2003. <<http://igm.univ-mlv.fr/~unitex/UnitexManual1.2.pdf>>
- Савъри 2000:** Savary, A. *Recensement et description des mots composés – méthodes et applications.* Autre [cs.OH]. Université de Marne la Vallée, 2000. Français. <<https://tel.archives-ouvertes.fr/tel-00003584/document>>
- Савъри 2005:** Savary, A. A formalism for the computational morphology of multi-word units. – *Archives of control sciences*, 2005, 15(3), pp.437-449.
- Савъри 2008:** Savary, A. Computational inflection of multi-word units a contrastive study of lexical approaches. – *Linguistic issues in language technology.* 2008, 1(1), pp. 1-53. <<http://journals.linguisticsociety.org/elanguage/lilt/article/download/lilt.1.2/71-214-1-PB.pdf>>
- Савъри 2009:** Savary, A. Multiflex: a multilingual finite-state tool for multi-word units. – *Lecture notes in computer science*, (5642), pp. 237-240.

- Сaг и др. 2002:** Sag, I., et al. Multiword expressions: a pain in the neck for NLP. – In: *Proceedings of the 3rd International Conference on intelligent text processing and computational linguistics (CICLing-2002)*. Berlin, Heidelberg: Springer, 2002, pp. 1-15. <<http://lingo.stanford.edu/pubs/WP-2001-03.pdf>>
- Стоянова 2012:** Стоянова, И. Автоматично разпознаване и тагиране на съставни лексикални единици в българския език. Дисертация за присъждане на образователната степен „доктор“, науч. рък. проф. Св. Коева. София: БАН, Институт за български език. Секция по компютърна лингвистика, 2012. <<http://dcl.bas.bg/sites/default/files/webfm/thesis-iva.pdf>>.
- Стоянова 2014:** Stoyanova, I. Automatic categorization of multiword expressions and named entities in Bulgarian. – In: *Proceedings of the First International Conference: Computational Linguistics in Bulgaria, 4 September 2014*. Sofia, Bulgaria, 2014, pp. 40-49.
- Тодорова 2009:** Тодорова, М. Лексикализирана граматика на български глаголни фразеологизми. – *Български език*, 2009, (3), с. 70–83.
- Тодорова и др. 2014:** Тодорова, М., Хр. Кукова и Св. Лесева. Семантично аотирани ресурси за български език – БулСемКор. – В: *Езикови ресурси и технологии за български език., състав. и науч. ред. Коева, Св. и Д. Благоева*, София: АИ „Проф. М. Дринов“, 2014, с. 80–104.
- Фишер и Кайл 1996:** Fischer, I., and M. Keil. Parsing decomposable idioms. – In: *Proceedings of the 16th Conference on Computational linguistics – Volkne 1*. Association for Computational Linguistics, 1996. 388–393. < <http://www.aclweb.org/anthology/C96-1066>>

Гл. ас. д-р Мария Тодорова
 Секция по компютърна лингвистика
 Институт за български език
 „Проф. Любомир Андрейчин“
 Българска академия на науките
 бул. „Шипченски проход“ 52, бл. 17,
 София 1113
 maria@dcl.bas.bg

Assist. Prof. Maria Todorova, PhD
 Department of Computational Linguistics
 Institute for Bulgarian Language
 “Prof. Lyubomir Andreychin”
 Bulgarian Academy of Sciences
 52 Shipchenski prohod blvd., building 17,
 Sofia 1113
 maria@dcl.bas.bg