

**ОПИСАНИЕ НА ОТНОШЕНИЯТА МЕЖДУ ГЛАГОЛНИТЕ
НЕСВОБОДНИ ФРАЗИ И ТЕХНИТЕ ДЕРИВАТИ
В ЛИНГВИСТИЧНИ РЕСУРСИ¹**

*Светлозара Лесева, Цветана Димитрова, Мария Тодорова,
Ивелина Стоянова*
*Секция по компютърна лингвистика,
Институт за български език, Българска академия на науките*

**DESCRIPTION OF VERBAL MULTIWORD EXPRESSIONS
AND THEIR DERIVATIVES IN LINGUISTIC RESOURCES**

*Svetlozara Leseva, Tsvetana Dimitrova, Maria Todorova,
Ivelina Stoyanova*
*Department of Computational Linguistics, Institute for Bulgarian
Language, Bulgarian Academy of Sciences*

The article aims at outlining a framework for a unified description of verbal multiword expressions and their derivatives in the lexical-semantic network WordNet. The framework is linguistically independent and is applicable to the systematic description of verbal multiword expressions in a wide variety of linguistic resources. Main principles of derivation are presented, as well as structural variations and semantic relations stemming from the derivation.

Key words: Computational linguistics, multiword expressions, WordNet, language resources, derivation

1. Въведение

Настоящата статия има за цел да представи рамка за унифицирано описание на глаголните несвободни фрази (ГНФ) и деривационно свързаните с тях единици в лексикално-семантичната мрежа Уър-

¹ Изследването е извършено в рамките на проекта „Обогатяване на лингвистични ресурси с деривационно свързани несвободни фрази“, изпълняван от Института за български език по договор за двустранно сътрудничество с Изследователския институт за изкуствен интелект към Румънската академия.

дНет, което да бъде приложено в Българския уърднет (БулНет). Рамката е езиково независима и е приложима и за систематичното описание на несвободните фрази в други типове лингвистични ресурси.

Несвободните фрази са съчетания от две или повече думи, чието общо значение не може да бъде пряко изведено от значенията на съставлящите ги компоненти (*давам на съд, отървавам кожата, бъркам в джоба*) и/или изискват специфично познание за съчетаемостта на елементите дори когато образуването им изглежда регулярно (срв. *провеждам изследване, извършвам изследване и провеждам интервю, но *извършвам интервю*). Описанието на ГНФ е актуален проблем както с оглед на автоматичното им разпознаване в свободен текст за целите на машинния превод, извличането на информация, разбирането и генерирането на текст и други области на компютърната обработка на езика, така и от гледна точка на отразяването им в лексикографските ресурси, при което се наблюдава нееднородност и непоследователност дори в рамките на един източник.

Формализираното представяне на ГНФ като езикови единици, предложено в статията, е основано на езиково независимо описание на морфосинтактичните, семантичните и синтактичните им особености, като наред с това се отразяват и деривационните отношения, в които встъпват с други несвободни фрази и сложни думи, и семантиката на тези отношения.

Езиковите единици в УърдНет са организирани в синонимни множества, свързани помежду си чрез система от семантични, лексикални, деривационни и други релации. ГНФ се въвеждат в съществуващи или нови синонимни множества, като наред с това се дефинират деривационни и семантични релации между тях и други единици (Фелбаум 1998).

В изложението са представени използваните езикови ресурси, възприетата система за унифицирано описание на несвободните фрази, подходите за въвеждане и верификация на единиците. В последната част са обобщени постигнатите резултати и са очертани някои насоки за бъдеща работа.

2. Представяне на несвободните фрази в лингвистични ресурси

В лексикографската практика се прилагат различни подходи за представяне на ГНФ и тяхното значение и характеристики, както и на съчетаемостта им. Обикновено те са включвани като част от описанието на един от съставлящите ги компоненти, но липсва последовател-

ност в това отношение (напр. в Речника на българския език² може да са представени в речниковата статия на глаголния или на именния компонент). В УърдНет ГНФ са включени като отделни литерали (синоними) в синонимно множество с подходящо значение (напр. {*участвам:2; вземам участие:1; взема участие:1*}³).

Според някои автори и в някои лексикални ресурси ГНФ са експлицитно отразени като отделни единици (Вилавиченцо и др. 2004; Коритковска 2008; Грегоар 2010; Стоянова и др. 2016), тъй като съчетаемостта между компонентите им не е изцяло предсказуема, а значението невинаги може да се изведе от значението на отделните им елементи. Наред с това често несвободната фраза има различна (идеосинкретична) аргументна реализация. Алтернативен подход е възприет в рамките на VALLEX, където за ГНФ с десемантизиран глагол описание на съчетаемостта е включено във валентните рамки както на глаголния, така и на именния елемент (Кетнерова и др. 2012).

Дефинирането на ГНФ като отделни речникови единици и компилирането на специализирани езикови ресурси с фокус върху тях позволява по-детайлно описание на морфосинтактичните им особености и ограничения, аргументната структура и семантиката им, семантичните им отношения с други единици, включително синонимията с други глаголи и несвободни фрази, както и деривационните отношения, в които встъпват. Спецификата на значението и характеристиките на ГНФ с оглед на компютърната обработка на езика налагат последователност при въвеждането им в лексикалните ресурси и създаването на унифицирана рамка за тяхното ефективно представяне.

За целите на настоящото изследване е използван следният набор от езикови ресурси:

(а) лексикално-семантичната мрежа БулНет⁴, съдържаща 3874 ГНФ;

(б) корпус с аотирани глаголни несвободни фрази, разработен в рамките на проекта PARSEME⁵, с обем 21 599 изречения и 480 413 думи и включващ 6721 аотирани ГНФ;

(в) електронна колекция от над 4000 ГНФ и техни деривати (Барбу Митителу и Лесева 2018).

² <http://ibl.bas.bg/rbe/> (15.01.2019)

³ В примерите от УърдНет са представени основните членове на дадено синонимно множество, като е посочен и уникалният номер на значението, напр. *участвам:2* показва, че това е второто поредно значение на думата *участвам*.

⁴ <http://dcl.bas.bg/bulnet/> (15.01.2019)

⁵ <https://gitlab.com/parseme/sharedtask-data/tree/master/1.1/BG> (15.01.2019)

Единиците са проверени автоматично, като са отстранени повторенията и в резултат е получена база от 7938 уникални ГНФ. Извършени са редица проверки, свързани с последователността и пълнотата на данните. Автоматично са генерирани и допълнени (където липсват) видовете съответствия на ГНФ от различните източници: например *притисвам в ъгъла* е генерирано автоматично като несвършен корелат на *притисна в ъгъла*, което фигурира в част от използваните ресурси.

Предвид организацията на данните в УърдНет под формата на синонимни множества е предприето групирането на несвободните фрази със синонимно значение; така например *поддържам отношения*, *поддържам контакт* и *поддържам връзки* принадлежат към едно общо множество. Предприето е и обединяване на различните варианти на несвободните фрази: с умалителна форма на съществителното (напр. *отървавам кожата* – *отървавам кожичката*); с модификатор, който може да се изпусне (напр. *пера мръсни пари* – *пера пари*); с членувана/нечленувана форма на съществително (напр. *попадам в клопка* – *попадам в клопката*) и др.

3. Типология и характеристики на ГНФ

Възприетата типология на ГНФ цели да обхване свойствата, определящи както разпознаването и описанието им, така и начините, по които образуват деривати. Типологията заимства класификацията на ГНФ, възприета в рамките на проекта PARSEME (Савари и др. 2017, 2018). В нея те се класифицират най-общо в четири големи класа:

(а) глаголни фразеологизми (ГФ), или идиоми (*verb idioms*, VID), с неразложимо значение (напр. *гушвам букета*, *обирам си крушите*);

(б) глаголни несвободни фрази с десемантизиран глагол (ДГНФ) (*light verb constructions*, LVC), които се разделят на два подкласа: същински (LVC.full; напр. *имам възможност*) и каузативни (LVC.cause; напр. *давам възможност*);

(в) глаголи със задължителни частици *се* или *си* (*inherently reflexive verbs*, IRV, напр. *усмихвам се*, *спомням си*);

(г) глаголни конструкции, в които съчетанието на глагол с определен предлог носи специфично значение (*inherently adpositional verbs*, IAV, напр. *заставам зад* = *подкрепям*).

В настоящата работа се съсредоточаваме върху първите два класа, тъй като именно те представляват интерес от гледна точка на съче-

таемостта между пълнозначни думи (глаголни и именни конституенти) и процесите на деривация.

Глаголните фразеологизми (ГФ) са същински неразложими несвободни фрази с неделимо значение. Някои от тези фрази може да имат както буквално, така и идиоматично значение, което да е неизводимо при съчетаване на значенията на конституентите (напр. *отварям си очите*), като често се отличават и с метафоричност (напр. *тегля чертата*; *вдигам гълъбите*). В структурно отношение се характеризират със сравнително фиксирана конституентна структура (напр. *отварям си/му/й/им... очите*, но не **отварям си очи*, **отварям си окото*), както и с ограничения при присъединяването на модификатори (*тегля една дебела черта*, но *вдигам *(тази/тежка) котва*).

ГНФ с десемантизиран глагол (ДГНФ) са конструкции със сложна предикация, изразена чрез съчетание от глагол и друг елемент, най-често именна или предложна фраза. Глаголът принадлежи към ограничено множество от глаголи с абстрактно значение, които изразяват аспектиалност, насоченост и под. (напр. *правя*; *нанасям/нанеса*; *оказвам/окажа*), докато вторият конституент (комплемент) представлява семантичен предикат, изразяващ абстрактна същност със събитийна семантика, като много често това е отглаголно съществително, производно на автосемантичен по отношение на съчетанието глагол (напр. *нанасям/нанеса удар – удрям*) (Коева и др. 2016; Винце и др. 2013). Комплементът може да бъде съществително име (NP; напр. *вземам решение*, *склучвам договор*), предложна фраза (PP; напр. *подлагам на мъчения*, *вземам под контрол*), прилагателно име (AP; напр. *правя луд*, *одирам жив*); наречие (AdvP; напр. *вземам предвид*).

Съчетаемостта между компонентите не е изцяло предсказуема (*давам прием*, но *правя купон*), а често е и езиково специфична (напр. български *хвърлям поглед* – английски *take a look*, буквално 'вземам поглед'). Различните глаголи често предполагат различна реализация на аргументите (напр. *вземам интервю от някого*, но *провеждам интервю с някого*, вж. Коритковска 2008).

Грамматичните характеристики на конституентите също не са изцяло предсказуеми. Именният елемент в някои конструкции може да се изменя по форма (напр. *взех интервюто*, *взеха интервютата*), а в други е неизменяем (напр. *взех участие*, но не и **взех участието*, **взех участия*). В редица случаи именният компонент може да подлежи на модификация (напр. *взех важно решение*).

И при глаголните фразеологизми, и при ДГНФ в общия случай е възможно вмъкването на допълнителни елементи, което произтича от

граматичното оформяне на изказването (напр. въпросителната частица *ли*, спомагателни глаголи в сложни глаголни форми) и синтактичната реализация на аргументите (местоименни клитики), напр. *Дали ли са им вече зелена светлина?*, *Взели ли са му вече интервю?*, както и на адвербиални фрази (напр. *Обирай си веднага крушите!*, *Нанасям бързо удара*), като тази възможност се подчинява на редица ограничения (напр. **Обирай си от къщата крушите!*), които за момента остават извън обекта на изследването.

4. Система за унифицирано описание на несвободните фрази в лингвистични ресурси

Системата се основава на концепцията за лексикографско описание на несвободни фрази, предложено у Коева и др. (2016), като включва и допълнителни елементи и модификации (вж. Пример 1). Описанието съдържа следните признаци:

(1) **Групов идентификатор на ГНФ:** ГНФ се обединяват в множества от единици, които се характеризират с общо лексикално значение, но се различават по лексикално-граматични характеристики (като глаголният вид) или по начина на глаголното действие (взема се предвид само общорезултативното значение). Груповият идентификатор представлява най-слабо маркираната в граматично отношение единица, например *пъхам зад решетките* е групов идентификатор на несвободните фрази {*пъхам зад решетките*, *пъхам зад решетките*, *пъхна зад решетките*}.

(2) **Основна форма на ГНФ:** представлява най-немаркираната форма на конкретната несвободна фраза.

(3) **Глаголна лема:** основната форма на глаголният конституент на фразата.

(4) **Вид на ГНФ:** взема се под внимание глаголният вид на опората на ГНФ, който предопределя глаголният вид на цялата фраза.

(5) **Синтактична структура:** повечето ГНФ спадат към някоя от следните основни структури:

- глагол + пряко допълнение: V + NP_обект: *вземам решение, давам зелена светлина;*
- субект + глагол: NP_субект + V: *звездата ми изгрява, метлата играе;*
- глагол + предложна фраза: V + PP: *подлагам на мъчения, поемам в свои ръце;*
- глагол + адективна фраза (сказуемно определение): V + AP: *правя луд, одирам жив;*

- глагол + адвербиална фраза: V + AdvP: *вземам предвид, чакам наготово.*

(6) Характеристики на опората: проявяват се със стойностите пълна парадигма (full) и ограничена парадигма (fixed). При ограничен парадигма се посочва какви са ограниченията, като обобщенията ще се обогатяват и прецизират въз основа на емпиричните данни.

(7) Характеристики на зависимите части: проявяват се със стойностите пълна парадигма (full) и ограничена парадигма (fixed): във втория случай се посочва какви са ограниченията. При зависима част предложна фраза се посочва и предлогът.

(8) Допустимост на модификатори към зависимата част: описва се със стойностите „да“/„не“ (Yes/No) за всяка зависима част и съответно вид (лексикална/фразова категория) на модификатора; други ограничения (Other), които се конкретизират.

(9) Допустимост за вмъкване на външни елементи между глагола и другите компоненти: описва се със стойностите „да“/„не“ (Yes/No) и съответния тип външен елемент: а) от общ вид (General): адвербиални фрази, частицата *ли*, кратки форми на личните и на притежателните местоимения, спомагателния глагол *съм* в сложни глаголни форми (регулярни елементи, които се допускат от вида и структурата на фразата); б) с други ограничения (Other), които се конкретизират; в) не допуска (None).

(10) Наличие на аргументни или спецификаторни позиции, които може да се запълнят в контекст: описва се със стойностите „да“/„не“ (Yes/No) и съответния тип аргументи – пряко допълнение (DO), непряко допълнение (IO), притежателен модификатор/притежателна PP (POSS).

(11) Възможност за словоредни изменения: дефинира се списък от регулярни словоредни варианти въз основа на общите особености на вътрешната синтактична структура (например за V + NP има възможни варианти „V NP“, „NP V“). В схемата за описание се включват само наблюдавани ограничения върху регулярните словоредни варианти.

(12) Възможност за семантични варианти и синоними: включват се синоними и варианти като описаните в т. 2, напр. *отървавам кожата, отърва кожата, отървавам кожичката, отърва кожичката, отървавам главата, отърва главата*. Предвижда се вариантите и синонимите (еднословни и несвободни фрази) да се извличат отчасти автоматично на базата на съществуващи вече ресурси с подобна информация.

(13) **Тип на несвободната фраза:** определя се според класификацията на несвободните фрази, възприета в проекта PARSEME (Савари и др. 2018). Типовете са приписани на несвободните фрази, анотирани в корпуса по PARSEME, както и на други ГНФ, които присъстват в предишни разработвани ресурси: Речник на несвободни фрази (Коева и др. 2016) и списъци с конструкции с десемантизирани глаголи (Стоянова и др. 2016).

(14) **Флективен тип на несвободната фраза:** флективният тип е обобщено означение на всички характеристики на несвободната фраза, които предопределят формоизменението ѝ: синтактична структура на фразата, флективен тип (спрежение) на глагола, флективен тип на изменяемите конституенти, ограничения върху парадигмата, синтактични вариации, възможни модификации на конституенти и др. Например към един и същ флективен тип спадат ГНФ *давам назад* и *вземам предвид* (глагол от несвършен вид, 3-то спрежение, с пълна парадигма, неизменяем подчинен конституент, без модификатори).

(15) **Възможни деривати:** За всяка ГНФ се дефинират възможните ѝ деривати (към момента вниманието е съсредоточено върху глаголно-именната деривация). Синтактичните особености на номинализираните фрази произтичат от синтактичните особености на типа фраза и от морфологичните свойства на опората, а в някои случаи – и на други елементи. Деривацията се базира на лингвистични обобщения, които дефинират основните закономерности, свързани с морфологичната форма, синтактичната реализация и граматичните ограничения, налагани върху главните и подчинените части (вж. т. 5). При наличието на граматични изменения или налагането на ограничения те се отбелязват в съответното поле (характеристики на опората; характеристики на зависимите части), например *разбивам (нечие) сърце* > *разбивач на сърца* (мн. ч., нечл.).

Пример 1. Описание на ГНФ

1	Групов идентификатор:	<i>служа на двама господари</i>
2	Лема на ГНФ:	<i>служа на двама господари</i>
3	Глаголна лема:	<i>Служа</i>
4	Вид:	<i>IMPERF</i>
5	Синтактична структура:	V PP; P = на; NP = B N
6	Парадигма на опората:	V-full
7	Парадигма на подчинената част:	Fixed

8	Модификатори:	No
9	Външни елементи:	General
10	Незаети позиции:	None
11	Словоредни ограничения:	0 (Регулярни варианти: V PP; PP V)
12	Синоними:	<i>лицемерничка; лицемеря;</i> <i>двуличен съм;</i> <i>лицемерен съм</i>
13	Тип на ГНФ (PARSEME):	VID
14	Флективен тип:	V IMPERF 2XY
15	Възможни деривати:	<i>служене на двама господари (event);</i> <i>слуга на двама господари (agent);</i> <i>служба на двама господари (event/state)</i>

5. Типология и описание на деривационните модели за несвободните фрази и съответстващите им семантични релации

Потенциалът на несвободните фрази да образуват нови несвободни фрази е слабо проучен проблем в българската езиковедска литература. Засяган е предимно въпросът за произвеждане на универбати (*спускам с парашут > парашутист*, Благоева 2011) или сложни думи от несвободни фрази или колокации (*вали сняг > снеговалеж*, Осенова и Симов 2015). Деривацията на несвободните фрази се изразява в произвеждане на нови несвободни фрази от вече съществуващи с помощта на словообразователни механизми, формообразователни средства и морфосинтактични трансформации (Лесева 2017).

Изучаването на деривацията при глаголните несвободни фрази е свързано с описанието на отношения между семантично свързани ГНФ и именни (*свързвам двата края > свързване на двата края*), адективни (*въоръжа се до зъби < въоръжен до зъби*) или други несвободни фрази, като между поне един от компонентите им има деривационно отношение. В настоящия етап от изследването фокусът е върху глаголно-именната (или именно-глаголната) деривация.

Описанието на деривационните отношения се основава на формално представяне на свойствата на ГНФ и свойствата на деривационно свързаната несвободна фраза⁶.

5.1. Основни закономерности

Наблюдават се следните зависимости:

(1) В голяма част от типовете деривация синтактичната опора в структурата на именната фраза е дериват на опората на ГНФ: *склучвам договор (с някого) > склучване на договор (с някого); служи на двама господари > слуга на двама господари;*

(2) Прекият обект в структурата на ГНФ се реализира в именната фраза като предложна фраза, въвеждана от предлога *на* или от предлог, специфициран от именния компонент на ГНФ: *радвам окото > радост за окото; търся силни усещания > търсач на силни усещания;*

(3) Субектният елемент в структурата на ГНФ се реализира в именната фраза като предложна фраза, въвеждана от предлога *на*: *(нечия) звезда изгрява > изгряването на (нечия) звезда;*

(4) Изискваните от изходната ГНФ предложни, адективни и адвербиални групи се реализират без промяна в именната фраза: *хвърлям прах в очите > хвърляне на прах в очите; влизам бос в огъня > влизане бос в огъня;*

(5) Незапълнени аргументни позиции (Коева и др. 2016) в ГНФ се реализират в именната фраза, както следва:

(а) незапълнена прякообектна позиция в ГНФ се реализира в именната фраза като модификатор или предложна група с притежателно значение: *пъхвам (някого) зад решетките > (нечие) пъхване/пъхването (на някого) зад решетките* (вж. също у Ницолова 2005: 74);

(б) незапълнена непрякообектна позиция в ГНФ се реализира в именната фраза като предложна група: *чета конско (на някого) > четене на конско (на някого); играя си на котка и мишка (с някого) > игра на котка и мишка (с някого);*

(в) Незапълнени модификаторни (притежателни) позиции (Коева и др. 2016) в ГНФ се реализират като модификатори и в именната фраза (модификатор или предложна група с притежателно значение): *бъркам/бъркане в (нечий) джоб/в джоба (на някого);*

Съществуват и други, по-слабо честотни структурни типове деривация:

⁶ В част от случаите ГНФ е не изходна фраза за деривацията, а резултатът от нея. Тук разглеждаме по-общия случай, но изводите са приложими и в обратната посока.

(7) Субектът или обектът в ГНФ се реализира в именната несвободна фраза не като предложна фраза, а като адекватен модификатор: при фраза NP_subject + V: *икономиката колабира > икономически колапс*; при V + NP_object: *инжектирам финанси > финансова инжекция* (Костова 2012);

(8) Адвербиалът в ГНФ съответства на адекватен модификатор в именната фраза: *мисля трезво > трезво мислене, трезва мисъл*.

5.2. Типология според вида на семантичната релация

Наблюдават се следните типове семантични отношения между ГНФ и деривационно свързани именни фрази:

(1) Събитийна/стативна деривация – деривационно свързаната именна фраза означава действието, процеса, събитието или състоянието, обозначено от ГНФ: *викам духове > викане на духове; играя карти > игра на карти; лежа на гърба (на някого) > лежане на гърба (на някого)*;

(2) Агентивна деривация – деривационно свързаната именна фраза означава вършител на обозначаващото от ГНФ действие, дейност или под.: *гълтам огън > гълтач на огньовете; викам духове > викач на духове; играя карти > играч на карти; служа на двама господари > слуга на двама господари*;

(3) Резултативна деривация – деривационно свързаната именна фраза означава резултат от обозначаващото от ГНФ действие, дейност или под.: *събирам куцо и сакато – сбирка на куцо и сакато*;

(4) Локативна деривация – деривационно свързаната именна фраза означава мястото на извършване на дадена дейност (действие) или протичане на даден процес, описван от ГНФ: *кова закони > ковачница на закони; пера пари > перачница на пари*;

(5) Инструментална деривация – деривационно свързаната именна фраза е инструмент на дейността (действието), означена от ГНФ: *хващам бримки > хващач на бримки*.

5.3. Структурни варианти на ГНФ с причастия

Наред с деривацията, при която глаголната опора на ГНФ съответства на (в общия случай) деривационно свързано съществително в именната несвободна фраза, се наблюдават и възможности за образуване на несвободни фрази с причастие на глаголната опора, които тук съвсем условно се наричат ‘структурни варианти’.

Възможността за образуването на структурни варианти с причастия от ГНФ зависи от изходната синтактична структура и свойствата

на глагола (вид, транзитивност). По-често наблюдаваните структурни варианти с причастие на глаголната опора включват:

(А) Варианти, образувани от ГНФ със структура V + NP_object и страдателно причастие на глагола – опорността се прехвърля върху именния елемент и се получава именна несвободна фраза с причастие модификатор: *взема решение > взето решение*;

(Б) Варианти, образувани от ГНФ със структура V + NP_object и деятелно причастие на глагола – причастиято става опора в адективна фраза: *взема решение > взелият решението (човек), вземащият решението (човек)*.

(В) Варианти, образувани от ГНФ със структура V + PP/AdvP/AP с незаета позиция на прекия обект – причастиято става опора в адективна фраза. Вариантите със сегашно деятелно причастие могат да модифицират именни опори, коиндексирани с контекстовия субект, напр. *Човекът удря (някого) под кръста > удрящият го под кръста човек*; вариантите със страдателно причастие могат да модифицират именни опори, коиндексирани с контекстовия обект, напр. *ударя човека под кръста > удареният под кръста човек*;

(Г) Варианти, образувани от ГНФ със структура NP_subject + V и деятелно причастие – образува се именна фраза с опора субектния елемент на ГНФ, а причастиято става негов модификатор: *(нечия) звезда изгрява > изгряваща звезда, (нечия) кръв кипне > кипнала кръв, (не-чии) клепачи натежат > натежали клепачи*.

Далеч не всички производни несвободни фрази или структурни варианти са лексикализирани; голяма част от тях представляват оказионализми (напр. *сбирка на куцо и сакато*) или имат употреба като контекстови варианти на ГНФ (срв. лексикализираните *хванат от гората* или *въоръжен до зъби и ударен под кръста*). От гледна точка на описанието и автоматичното откриване в езиков контекст на такива единици е важно всички те да могат да бъдат разпознавани. В лексикалните ресурси обаче следва да се включват само тези от тях, които са лексикализирани.

6. Резултати и задачи за бъдеща работа

Описанието на несвободните фрази в съответствие с предложената система и разглеждането на семантичните и граматическите им характеристики, които предопределят наличието на специфични деривати, е дългосрочна задача.

Идентифицирането на възможни деривати (вж. 4, характеристика 15 от описанието) е извършено автоматично на базата на граматичес-

ките характеристики на изходната несвободна фраза и на съществуващите семантични релации, в които участва синонимното множество в УърдНет, съдържащо несвободната фраза. Например при глаголи от несвършен вид се предполага възможна деривация (ГНФ > именна несвободна фраза с опора отглаголно съществително име със събитийно значение), каквато не е дефинирана за глаголи от свършен вид.

Към момента е извършено автоматично описание на 7825 глаголни несвободни фрази, като към тях са добавени още 321 верифицирани несвободни фрази въз основа на допълнително намерените видови двойки. Общият брой групи е 3874. Описанието на характеристиките е извършено (полу)автоматично въз основа на наличната верифицирана вече информация в използваните ресурси, поради което не всички полета за описание са запълнени. Прилагат се и някои евристични подходи за попълване на описанието: подчинената част се специфицира с пълна парадигма, ако несвободната фраза е съчетание с десемантизиран глагол, а самата подчинена част е именна фраза (NP); или с ограничена парадигма, ако несвободната фраза е глаголен фразеологизъм или ако е съчетание с десемантизиран глагол и подчинената част е предложна фраза (в тези случаи неизменяемо е съществителното в предложната фраза) и др.

За някои от дериватите автоматично се генерира възможната форма въз основа на налични деривационни релации между глагола в ГНФ и думи в УърдНет: (а) изведени пряко – думите *викач*, *играч* са въведени в БулНет като деривационно свързани съответно с *викам*, *играя* с морфосемантична релация Агент, поради което е възможно образуването на двойката: *викам духове* > *викач на духове*; (б) изведени косвено – *слуга* присъства като деривационно свързано с релация Агент към *служувам:1*, което е в същия *verb_group* със *служва:4*, и по този начин можем да се допусне деривация от вида *служва на двама господари* > *слуга на двама господари*; (в) изведени въз основа на отдалечена асоциация – *хващач* присъства като Агент на *хващам*, но не и като Инструмент, но наставката *-ач* е характерна и за инструменти, поради което е възможно извеждането на двойката *хващам бримки* > *хващач на бримки*. Има и глаголи без налични деривативно свързани съществителни имена в БулНет (напр. глаголят *гълтам* е агентивен, но в УърдНет няма намерен еднословен еквивалент, който да служи като негов Агент), за които може да се приложи автоматично допълнително търсене в други езикови ресурси с подходящи деривационни модели.

Автоматично изведените деривати е необходимо да бъдат верифицирани ръчно. Попълването на липсваща информация също е задача

за бъдеща работа. Разширяването на данните ще допринесе за по-пълнен и по-детайлен обзор на типологията, както и за описанието на характеристиките на дериватите и на семантичните релации, основаващи се на деривационни отношения.

ЛИТЕРАТУРА

- Барбу Митителу, Лесева 2018:** Barbu Mititelu, V., Leseva, S. Derivation and Multiword Expressions. // *Multiword Expressions: Insights from a Multilingual Perspective*. Phraseology and Multiword Expressions Series. Berlin: Language Science Press, 2018, 215 – 246.
- Благоева 2011:** Благоева, Д. Дефразеологизацията като източник на лексикални и семантични иновации в съвременния български език. [Blagoeva, D. Defrazeologizatsiyata като iztochnik na leksikalni i semantichni inovatsii v savremenniya balgarski ezik.] // *Езиковедски изследвания в чест на проф. Сийка Спасова-Михайлова*. София: АИ „Проф. Марин Дринов“, 2011, 131 – 142.
- Вилавиченцо и др. 2004:** Villavicencio, A., Copestake, A., Waldron, B., and Lambeau, F. The lexical encoding of MWEs. // *Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing*. Barcelona, 2004, 80 – 87.
- Винце и др. 2013:** Vincze, V., Nagi T., I., Farkas, R. Identifying English and Hungarian light verb constructions: A contrastive approach. // *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013, 255–261. 15 January, 2019 <http://aclweb.org/anthology/P13-2046>.
- Грегоар 2010:** Gregoire, N. Dueme: a Dutch electronic lexicon of multiword expressions. // *Language Resources and Evaluation*, 2010, 44: 23 – 39.
- Кетнерова и др. 2012:** Kettnerova, V., Lopatkova, M., Bejcek, E. The Syntax-Semantics interface of Czech verbs in the Valency Lexicon. // *Proceedings of the 15th EURALEX International Congress*. Oslo, 2012, 434 – 443.
- Коева и др. 2016:** Коева, S., Stoyanova, I., Todorova, M., Leseva, S. Semi-automatic Compilation of the Dictionary of Bulgarian Multiword Expressions. // *Proceedings of the GLOBALEX 2016 Workshop: Lexicographic Resources for Human Language Technology*. LREC, 2016, 86 – 95.
- Коритковска 2008:** Коритковска, М. Съчетанията глагол + съществително като лексиколожки и лексикографски проблем. [Koritkovska,

- М. Sachetaniyata glagol + sashtestvitelno kato leksikolozhki i leksikografski problem.] // *Изследвания по фразеология, лексикология и лексикография (в памет на проф. д.ф.н. Кети Анкова-Ничева)*. София: АИ „Проф. Марин Дринов“, 2008, 227 – 232.
- Костова 2012:** Костова, Н. Глаголи с нови значения в българския език. [Kostova, N. Glagoli s novi znacheniya v balgarskiya ezik.] // *Славистиката в глобалния свят – предизвикателства и перспективи*. Благоевград: УИ „Неофит Рилски“, 2012, 15 – 22.
- Лесева 2017:** Лесева, С. Несвободни фрази и деривация. [Leseva, S. Nesvobodni frazi i derivatsiya.] // *Известия на Института за български език „Проф. Любомир Андейчин“*. София: Издателство на БАН „Проф. Марин Дринов“, 2017, 262 – 303.
- Ницолова 2005:** Ницолова, Р. За формалното означаване на семантичните роли в българските конструкции с отглаголни съществителни. [Nitsolova, R. Za formalното oznachavane na semantichnite roli v balgarskite konstruksii s otglagolni sashtestvitelni.] // *Аргументна структура. Проблеми на простото и сложното изречение*. София: СемаРШ, 2005, 69 – 79.
- Осенова, Симов 2015:** Osenova, P., Simov, K. Modeling lexicon-syntax interaction with catenae. // *Journal of Cognitive Science*, 2015, 16 (3): 287 – 322.
- Савари и др. 2017:** Savary, A., Ramisch, C., Cordeiro, S., Sangati, F., Vincze, V., QasemiZadeh, B., Candito, M., Cap, F., Giouli, V., Stoyanova, I., Doucet, A. The PARSEME shared task on automatic identification of verbal multiword expressions. // *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*. Valencia: Association for Computational Linguistics, 31 – 47. 15 January, 2019. <<http://www.aclweb.org/anthology/W17-1704>>.
- Савари и др. 2018:** Savary, A. (ed.). *Annotation guidelines of PARSEME shared task on automatic identification of verbal MWEs*. Edition 1.1. 15 January, 2019 <<http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.1/>>.
- Стоянова и др. 2016:** Stoyanova, I., Leseva, S., Todorova, M. Towards the Automatic identification of light verb constructions in Bulgarian. // *Proceedings of CLIB 2016. Proceedings of the International Conference Computational Linguistics in Bulgaria*. Sofia: Institute for Bulgarian Language, 2016, 28 – 37.
- Фелбаум 1998:** Fellbaum, C. (ed.). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.