

ПРИНЦИПИ НА ФЛЕКТИВНОТО ОПИСАНИЕ НА ГЛАГОЛНИТЕ СЪСТАВНИ ЛЕКСИКАЛНИ ЕДИНИЦИ В БЪЛГАРСКИ

СВЕТЛА КОЕВА

ИНСТИТУТ ЗА БЪЛГАРСКИ ЕЗИК „ПРОФ. ЛЮБОМИР АНДРЕЙЧИН“

ПРИ БЪЛГАРСКАТА АКАДЕМИЯ НА НАУКИТЕ

svetla@dcl.bas.bg

Статията представя общите принципи за флективно описание на глаголните съставни лексикални единици в български. Прилагат се два основни критерия: запазване на връзката с флективното описание на простите думи в българския език и представяне на минимално необходимата граматична информация, която осигурява единно описание на формите на съставните лексикални единици. Представен е накратко използваният формализъм за дефиниране на локални граматики, които се компилират в крайни преобразуватели за разпознаване и генериране на формите на съставните лексикални единици и прилежащата им граматична информация.

Ключови думи: флективни типове; глаголни съставни лексикални единици; глаголни фразеологизми; български език

PRINCIPLES OF THE INFLECTIONAL DESCRIPTION OF VERBAL MULTIWORD EXPRESSIONS IN BULGARIAN

SVETLA KOEVA

INSTITUTE FOR BULGARIAN LANGUAGE, BULGARIAN ACADEMY OF SCIENCES

svetla@dcl.bas.bg

This article presents the general principles for the inflectional description of verbal multiword expressions in Bulgarian. Two main criteria are applied: the maintenance of the relation to the inflectional description of simple words in Bulgarian and the presentation of the minimal grammatical information necessary for a standardised description of the forms of verbal multiword expressions. The paper also provides a brief outline of the formalism used to define local grammars, which are compiled into finite-state transducers that recognise and generate the forms of verbal multiword expressions, along with the associated grammatical information.

Keywords: inflectional types; verbal multiword expressions; verbal phraseologisms; Bulgarian

1. Въведение

Въпросът за автоматичната обработка на съставните лексикални единици, свързана с формалното описание на формите им, се поставя сравнително отскоро (Gross 1990; Silberstein 1993; Breidt et al. 1996; Gross 1997; Silberstein 2005; Savary 2005; Koeva 2006 и др.), тъй като е доста по-трудно да се опишат техните флективни типове в сравнение с флективните типове на думите. Предлагани са различни описания и дефиниции за съставни лексикални единици (Bally 1909; Виноградов/Vinogradov 1947: 28–33; Zgusta 1967; Benveniste 1974; Кювлиева-Мишайкова/Kyuvlieva-Mishaykova 1986: 10–16; Ничева/Nicheva 1987: 19; Gross 1990; Nunberg et al. 1994: 497–498; Sag et al. 2002: 190; Koeva 2006: 201; Baldwin, Kim 2010: 269–272 и др.), като, както се посочва в обобщение (Savary 2008: 2), фокусът е върху няколко основни компонента. Съставните лексикални единици (multiword expressions на английски) са:

- **многокомпонентни**: състоят се от две или повече (графични) думи;
- **идиоматични**: характеризират се със (някаква степен на) семантична некомпозиционалност¹;
- **устойчиви**²: референцията им е уникална.

Възприемаме следната компилирана дефиниция за съставни лексикални единици: последователност от думи (включително прекъснатата), която има уникална (и относително константна във времето) референция и се характеризира със (някаква степен на) семантична некомпозиционалност, съчетана в някои случаи с нетрадиционна за дадения език лексикална, морфологична, синтактична или прагматична употреба.

Целта на настоящата статия е да се разгледа формалното представяне на флективните типове³ на глаголните съставни лексикални единици (глаголните фразеологизми)⁴, част от активната лексика на българския език. Предмет на изследването са глаголните фразеологизми, включени в *Речник на активната фразеология в българския и полския език* (Благоева и др./Vlagoeva et al. 2022), които са класифицирани по флективни типове (или групи, в които лексикалните единици имат еднакво формообразуване).

Принципите на формалното описание на флективната морфология на съставните лексикални единици в българския език вече са описани като общ модел (Коева 2006), с повече детайли (Стоянова/Stoyanova 2012; Стоянова, Тодорова/Stoyanova, Todorova 2014; Тодорова/Todorova 2015)⁵ и достатъчно изчерпателно за практическа работа (Коева et al. 2016). Формалното представяне на флективните типове на глаголните фразеологизми включва дефиниране на основната форма, формулиране на единен подход за описание, основаващ се на лексикалната и граматичната структура на езика, и приложение на избрания формализъм за генериране (разпознаване) на формите от парадигмата. Ще бъдат разгледани с повече подробности възприетите принципи при дефинирането на основната форма на глаголните фразеологизми, което предопределя и членовете на парадигмата им.

Разглежда се флективното описание само на глаголните фразеологизми, чиято структура съвпада със структурата на глаголните подчинителни словосъчетания: [*хващам [бика] [за рогата]*] и [*хващам [чантата] [за дръжката]*]]. Извън обсега на изследването остават фразеологизми със структура на самостоятелни прости изречения (*очите ми шарят*), на подчинени изречения (*без да ми мигне окото*), на сложни съчинени изречения (*от едното ухо влиза, от другото излиза*) и на сложни съставни изречения (*режа клона, на който седя*). Макар че се вписват в предлаганото описание, към момента извън него остават и постоянно рефлексивните, постоянно реципрочните, постоянно акузативните, постоянно дативните и постоянно дативно-рефлексивните глаголи: *смея се, въобразявам си, състезавам се, мързи ме, хрумна ми, гади ми се*.

2. Характеристики на глаголните съставни лексикални единици, които имат отношение към флективното им описание

Отдавна е забелязано, че (част от) съставните лексикални единици могат да се променят в зависимост от морфосинтактичния контекст: компонентите им могат да имат различни форми и различен словоред, а някои от тях могат да бъдат разделени от думи или словосъчетания⁶.

2.1. Формоизменение на компонентите на глаголните фразеологизми

Безспорно възможностите за **формоизменение**, наречени формална вариантност (Ничева и др./Nicheva et al. 1974: 27–28), на част или на всички компоненти от даден глаголен фразеологизъм са **част от флективното му описание**. Компонент на глаголният фразеологизъм може да е неизменяема дума, парадигмата му да е ограничена до една или няколко форми или да съвпада с парадигмата на думата извън фразеологизма. Особеностите и ограниченията във формообразуването (дефектни парадигми) са били предмет на описание (Кювлиева-Мишайкова/Kyuvlieva-Mishaykova 1986: 33; Ничева/Nicheva 1987: 77–78 и др.) от гледна точка на изчерпателното описание на характеристиките на българските фразеологизми.

Макар че една от характеристиките на глаголните съставни лексикални единици е фиксираността на формата на компонентите, всъщност при повечето глаголни фразеологизми глаголът подлежи на формоизменение, а в някои случаи и прилежащите му компоненти могат да се изменят. Доколкото във флективните типове на простите думи се включват само синтетичните форми – тъй като аналитичните форми са изводими с правила, ако флективните типове са дефинирани правилно, то подходът при флективното описание на глаголните фразеологизми е същият: включват се **само синтетичните форми на глаголите**. Особеното при съставните лексикални единици е, че броят на редовете на парадигмата на главната част може да се умножи по броя на редовете на парадигмата на зависимите изменяеми части: *влизам в релси, влизам в релсите; влизааш в релси, влизааш в релсите* и т.н.

2.2. Синтактична структура на глаголните фразеологизми и словоредни варианти на компонентите

Предлага се дефиниране на фразеологизмите, което се основава на схващането, че лексикалната опора лицензира зависимите си части, т.е. семантичните и синтактичните свойства на опората определят броя и типа на останалите конститuentи, с които тя може или трябва да се съчетава: *Редицата x...y...z (в която редът на елементите е без значение) образува верига, ако x лицензира y и z или ако x лицензира y и y лицензира x. Елементите на фразеологизма трябва да образуват верига* (O'Grady 1998: 284). Фразеологизмите се формират, когато определена верига от опори придобие фигуративно значение (O'Grady 1998: 289). Подходът се основава на разбирането, че фразеологизмите и словосъчетанията в даден език имат еднаква структура, или по-точно: фразеологизмите могат да се реализират в рамките на възможните за даден език синтактични структури на свободните словосъчетания. Това заключение е направено от редица учени (Fellbaum 1993: 274; Калдиева-Захаријева/Kaldieva-Zaharieva 2013: 120; Коева/Koeva 2024: 63–64 и др.). Може да се обобщи, че синтактичната структура на словосъчетанията (и изреченията), характерна за даден език, предопределя структурата на фразеологизмите в този език.

В български (най-общо, на базата на *Речник на активната фразеология в българския и полския език*) фразеологизмите с главна част глагол могат да съдържат: именна група, предложна група и съчетания от именна група и предложна група или две предложни групи (по-рядко се срещат и структури с наречие или с малко изречение) (Коева/Koeva 2024: 71–75)⁷. Именната група може да се изразява от отделно съществително име: *вода парада*, (рядко) от съчинително свързани съществителни: *бълвам змии и гущери*, или от главна част съществително име, разширено в препозиция с един или (рядко) повече модификатори: *разкривам истинското си лице*. Предложната група в рамките на глаголните фразеологизми обикновено се състои от предлог и съществително име, рядко от предлог и именна група: *настъпвам (някого) по мазола, лепя се като гербова марка*. Възможностите за реализация на фразеологизмите в структурно отношение са сравнително добре представени, доколкото е описана структурата на глаголните словосъчетания в българския език (Бъркалова/Barkalova 1997; Пенчев/Penchev 1998; Коева/Koeva 1999 и др.).

В описанието на флективните типове на глаголните фразеологизми синтактичната структура може да се кодира в основната форма на равнище синтактични зависимости (главна част – подчинена част, съчинени части) между компонентите на фразеологизма, ако подобно кодиране се окаже различително за отделни флективни типове. До момента подобна необходимост не е налице и от категориите и реда на компонентите може да се изведе информация за синтактичните отношения.

Макар че рядко се обръща внимание на словоредните варианти на глаголните фразеологизми (Ничева и др./Nicheva et al. 1974: 12; Кювлиева-Мишайкова/Kyuvlieva-Mishaykova 1986: 35), наблюденията показват, че ако фиксираният словоред не е характеристика на фразеологизма, глаголните фразеологизми с определена структура допускат същите словоредни промени като съответните свободни словосъчетания: *гледам като теле в железница* и *като теле в железница гледам* (*отивам към училището в центъра* и *към училището в центъра отивам*); *вдигам си чуковете* и *чуковете си вдигам* (*вземам си багажа* и *багажа си вземам*); *лягам си с кокошките* и *с кокошките си лягам* (*лягам си със съседката* и *със съседката си лягам*). **Описанието на словоредните комбинации** между компонентите на глаголните съставни лексикални единици също **не е част от флективното описание**, доколкото то може да се формализира с допълнителни синтактични правила, по принцип валидни за българската синтактична структура (както по отношение на свободните словоредни размествания, така и по отношение на фиксирания словоред).

2.3. Вмъкване на външни елементи между компонентите на глаголните фразеологизми

При съставните лексикални единици се наблюдава още една особеност, свързана с фиксирания и променливия словоред: възможността за вмъкване на елемент от изречението, който не е част от съставната лексикална единица. Тази възможност е отбелязана отдавна, като се посочва, че най-често това са прилагателни, числителни или местоимения като формални определения към именен компонент или наречия като формални обстоятелствени пояснения към глаголен компонент (Ничева и др./Nicheva et al. 1974: 23): *вземам си такава беля на главата, държа здраво фронта*.

Наблюдават се три възможности в български (Коева 2006: 211–212): не се допуска вмъкване на външен за фразеологизма елемент: *умря Марко*; допуска се вмъкването само на клитики: *коравата му глава*; допуска се вмъкването на различни елементи: *хвърлям му неочаквано обвинения в лицето*. Вмъкването на произволен елемент (включително клитика) не е част от флективното описание на съставните лексикални единици. Затова приетият подход е да се описват само ограниченията (за фиксиран словоред) и възможностите (за вмъкване на компонент извън състава на фразеологизма).

3. Дефиниране на основната форма на глаголните фразеологизми

Основната форма на фразеологизмите може да се разглежда или като каноничен представител на парадигмата, или като идентификатор, ако е абстрактна, езиково неправилна форма, състояща се от основните форми на компонентите (Savary 2008: 10)⁸, например *ближа подметка негов* вместо *ближа подметките му*. Приемаме, че основната форма е част от парадигмата (Коева et al. 2016: 87) и се състои от лемата (или от най-неутралната

фразеологизирана форма) на главната част, в случая на глагола, и от лемите (или от най-неутралните фразеологизирани форми) на останалите компоненти. Основната форма на простите (и сложните) думи е *максимално немаркираната, т.е. тази форма, при която няма морфемно изразени граматични значения*, с изключение на глаголите, при които формата за първо лице е предпочетена пред немаркираното трето лице (Куцаров/Kutsarov 2007: 20). Под най-неутрална форма имаме предвид най-немаркираната за граматични значения форма, която има реална употреба (Коева/Коева 2008: 25). Съответно най-неутралната фразеологизирана форма на даден компонент на глаголна съставна лексикална единица е формата, която има най-малко изразени граматични значения в сравнение с останалите реализирани форми от парадигмата (или е неизменяема). Например основната форма на глаголния фразеологизъм *отварям нова страница* се състои от лемите на глагола *отварям* и съществителното *страница*, тъй като са част от парадигмата на фразеологизма, и от нечленуваната форма за женски род единствено число *нова*, защото това е морфологично най-немаркираната форма на прилагателното *нов*, която се употребява във фразеологизма (в случая единствената).

3.1. Подбор на основната форма на глаголните фразеологизми в сравнение с фразеологичните речници

Разпространена практика във фразеологичните речници е фразеологизмите с глаголи, в чийто състав могат да влязат глаголи и от свършен, и от несвършен вид, да се представят заедно (Ничева и др./Nicheva et al. 1974; Maldžieva, Wójtowicz 1994; Благоева и др./Blagoeva et al. 2022 и др.). Тъй като образуването на нови глаголи от свършен вид с представки, както и вторичната имперфективация са процеси, които засягат в много случаи и глаголните фразеологизми, а разликата в глаголния вид е свързана с разлика в парадигмите, при флективното описание е възприет следният принцип:

Фразеологизми, в чийто състав присъства глагол от несвършен или от свършен вид, се разглеждат независимо един от друг: *отварям нова страница* и *отворя нова страница*; *минавам всяка граница* и *мина всяка граница*; *преминавам всяка граница* и *премина всяка граница*.

3.2. Факултативни елементи в състава на глаголните фразеологизми

В някои фразеологизми присъстват факултативни елементи, които могат да бъдат или да не бъдат част от фразеологизма. Възможен подход е те да бъдат маркирани в лемата, например със скоби, както е в *Речник на активната фразеология в българския и полския език* (Благоева и др./Blagoeva et al. 2022: 10): *виря си нос (до небето)*.

Тъй като при компютърната обработка въпросът за икономията на ресурсите не стои по същия начин, както при традиционните речници, възприетият подход при флективното описание е глаголният фразеологизъм

вирия си носа (до небето) да бъде развит в две единици: *вирия си носа* и *вирия си носа до небето* (разбира се, връзката между двата фразеологизма може да се отрази).

Трябва да се направи разлика с външните елементи, които е допустимо да се вмъкват в някои глаголни фразеологизми, но не са част от състава на фразеологизма: *вирия си нахално носа до небето*. Както и в други речници (Ничева и др./Nicheva et al. 1974: 24), тези елементи не се включват в състава на основната форма.

3.3. Задължителни разширения към глаголните фразеологизми

Възможно е глаголните съставни лексикални единици да присъединяват задължителни разширения, които обаче не са лексикализирани като част от фразеологизма: *вкарвам някого в релсите* – *вкарвам учениците в релсите*, *вкарвам непослушните ученици в релсите*, *вкарвам непослушните ученици от десети клас в релсите*.

Случаят е аналогичен със семантично и синтактично задължителните части в глаголното словосъчетание (и в изречението), които се определят от характера на описваната от предиката ситуация и се класифицират с термина *аргументи*: *почитам някого*; *вярвам в някого*; *благодаря на някого* и т.н. Следователно семантично (и синтактично) задължителните разширения, които не са лексикализирани като част от глаголните фразеологизми, **не са част** от флективното им описание и от основната им форма; те са предмет на валентно описание на семантично и синтактично равнище.

Най-общо случаите на задължителни разширения към глаголните съставни лексикални единици могат да се опишат по следния начин: задължителна фразеологично нелексикализирана именна група, която се комбинира с фразеологизирана предложна група: *изхвърлям някого на улицата*; *замитам нещо под килима*; фразеологично нелексикализирана предложна група, която се комбинира с фразеологизирана предложна група: *махам с опашка пред някого*; *радвам се като дете на нещо*. Лемите на изброените глаголни фразеологизми са, както следва: *изхвърлям на улицата*; *замитам под килима*; *махам с опашка*; *радвам се като дете*.

В речниците към основната форма на фразеологизмите от този тип е прието да се посочва, че те функционират със задължително разширение: *някого*, *нещо*, *на някого* (Ничева и др./Nicheva et al. 1974; Maldžieva, Wójtowicz 1994; Благоева и др./Blagoeva et al. 2022 и др.). Възприетият подход при флективното описание е същият: към основната форма се описва семантично и синтактично задължителният компонент, но не се включва в парадигматичното описание: *изхвърлям на улицата (някого)*; *изхвърляш на улицата*; *изхвърля на улицата* и т.н.

Това означава, че извън флективното описание остават и формите, при които задължителните разширения се изразяват с кратки форми на личните местоимения: *изхвърлям го на улицата*; *замитам го под килима*; *радвам му*

се като дете. Трябва да се посочи, че има фразеологизми с постоянно акузативни, постоянно дативни и постоянно дативно-рефлексивни глаголи, в чийто състав компоненти са и кратките форми на личните местоимения: *мед ми капе на сърцето, дреме ми на шапката, вдига ми се кръвното*, както и (по-рядко) някои лични глаголи: *действам му като червено на бик*.

3.4. Изразяване на именни притежатели в структурата на глаголните фразеологизми

Включването на именен притежател в състава на глаголните фразеологизми в някои случаи е задължително както семантично, така и синтактично: *забива ми се в сърцето, изскочи ми от главата*. Изразяването на задължителен именен притежател към конституент в глаголния фразеологизъм може да стане по три начина (съвпадащи с възможностите при свободна употреба): с предложна група (*взех акъла на хората*), с пълна или с кратка форма на притежателно местоимение (*взех техния акъл, взех им акъла*). Когато именният притежател е изразен с предложна група, в неговия състав може да бъде включено произволно съществително или именна група, които обозначават лице, като същото се отнася и за местоименията – могат да са анафори на произволно съществително или именна група от контекста, които обозначават лице. При рефлексивен именен притежател обичайната възможност за изразяване е чрез кратка форма на рефлексивното притежателно местоимение: *пъхам си носа (някъде)*.

В речниците (до голяма степен последователно) фразеологизми от този тип се оформят без именния притежател, като се посочва, че функционират задължително с такъв елемент: *лежа на нечий гръб и лежа на гърба на някого*; *плувам в нечи води и плувам във водите на някого* (Ничева и др./Nicheva et al. 1974: 25); *попадам в лапите на някого, забива се в сърцето ми* (Благоева и др./Blagoeva et al. 2022); *по вкуса съм (нечий, на някого); чета мислите (нечи, на някого)* (Maldžieva, Wójtowicz 1994).

При кодирането на тази информация в лемата следваме принципа, възприет в българската граматика и лексикография, по отношение на постоянно акузативните, постоянно дативните и постоянно дативно-рефлексивните глаголи, където вътрешен аргумент може да се изрази с произволно съществително или именна група, които обозначават лице, но се представят чрез кратка винителна или дателна клитика: *боли го Петър главата, домъчня им на майката и бащата за малкото дете, свиди ѝ се на свекървата да даде свещника*. Аналогията не е пълна, защото при посочените класове глаголи в трето лице личната местоименна клитика не може да се замени със съществително или предложна група, а само да се дублира с тях.

Тъй като в разглежданите случаи притежателят е едновременно произволен и задължителен, в лемата посочваме най-разпространената реализация на притежателя – кратката форма на притежателното местоимение (*вземам му акъла*). Избира се формата за трето лице мъжки род, тъй като при

глаголните фразеологизми основната форма на личните глаголи е в първо лице и представянето на притежателното местоимение трябва да е в различно лице: **вземам ми акъла*. Останалите две конструкции са изводими с правила – конструкцията с предложна група не е често срещана, а конструкцията с пълна форма на притежателното местоимение се среща много рядко. Ако притежателят е рефлексивен, в общия случай е възможна само една употреба – с рефлексивна притежателна клитика: *посипвам си главата с пепел*. Срещат се и фразеологизми, в които се употребяват само една или две от трите възможности, в такъв случай основната форма съдържа местоимението: *вземам нечия страна*.

Формално редиците глагол – клитика – съществително при глаголните фразеологизми с едно разширение именна група и с две разширения – именна и предложна група, може да съвпадат: *натривам му носа* и *погаждам му номер*, но в първия пример клитиката е притежателна (*натривам неговия нос*), а във втория – лична дателна (**погаждам неговия номер*). Притежателната клитика е модификатор в рамките на именната група, която е конституент в глаголния фразеологизъм, а личната дателна клитика е възможна реализация на семантично и синтактично задължителен елемент, който не е част от глаголния фразеологизъм. В подобни случаи има основание да се въведе терминът синтактичен апроксимат, като явлението може да се наблюдава както в рамките на един език, така и междуетниково (при сравнение на български и **някои от славянските езици**)⁹.

В обобщение, основната форма на глаголните съставни единици включва всички компоненти, които са задължителни съставни части на фразеологизма на лексикално равнище, включително задължителен притежател: *скроявам капан/скроя капан, бълвам змии и гущери, минавам всяка граница/мина всяка граница, ближа подметките му, глътвам езика си/гълтна езика си, спя като пън, удрям с един куришум два заека/ударя с един куришум два заека, ловя риба в мътна вода* и др. В основната форма кратките притежателни форми се разполагат след съществителното, към което се отнасят, макар че най-разпространената употреба е да са в позиция след глагола. Причината за такъв подход при формалното представяне е по-лесното кодиране на празните позиции, ако са в началото или края на фразеологичната единица, когато това не рефлектира в образуването на несъществуващи лексикални единици.

4. Формализъм за представяне на българските глаголни фразеологизми

Всеки речник на флективните типове се основава на някакъв тип формализъм. Използвани са разнообразни формализми, които са обобщени по следния начин (Lion-Bouton et al. 2023: 122): **формализми, базирани на регулярни изрази или подобни операции** (Silberztein 1993; Breidt et al. 1996; Коева/Коева 1998; Chrobot et al. 1999; Oflazer et al. 2004; Silberztein 2005; Коева 2006; Sailer, Trawiński 2006, Паскалева/Paskaleva 2007; Стоянова/Sto-

уанова 2012; Стоянова, Тодорова/Stoyanova, Todorova 2014; Тодорова/Todorova 2015 и др.), в които компонентите на съставните лексикални единици са изброени в реда, в който могат да се появяват, а прекъсваемостта най-често се обозначава със специални символи, налагащи ограничения върху възможните вмъквания (или чрез ограничаване на броя на вмъкванията, или чрез ограничаване на думите, които могат да бъдат вмъкнати); **формализми, базирани на фразови граматика** (Grégoire 2010; Przepiórkowski et al. 2017; Savary et al. 2018 и др.), в които съставните лексикални единици са част от граматични правила, а прекъсванията се обозначават според възможностите на формализма; **реляционни бази от данни** (Vondříčka 2019) и др.

Формализмите, които се използват за описание на формообразуването на простите и сложните думи, могат да се приложат и при описанието на съставните лексикални единици, макар че поради особеностите на структурата на последните описанието е много по-комплексно. Степента на сложност се обосновава и от предназначението на формалното описание: ако целта е автоматично разпознаване на съставните лексикални единици в текст, описанието може да е на по-елементарно равнище; ако целта е едновременно разпознаване и генериране на формите на съставните лексикални единици спрямо дадени основни форми, то тогава формалното описание трябва да включва всички комбинации при формообразуването на компонентите на съставните лексикални единици, възможностите за словоредни варианти и за вмъкване на компоненти, външни за структурата на съставните лексикални единици.

Възприема се подходът за унифицирано представяне на простите (сложните) и съставните лексикални единици (Courtois, Silberstein 1990; Koeva 2006; Savary 2008 и др.). Използваният формализъм се основава на теорията за крайните автомати, по-конкретно на локалните граматика, формализирани формообразуването на морфологично и морфосинтактично равнище под формата на преобразуватели с крайни състояния, които генерират всички морфологични форми на думите, включени в речника. За улеснение се използват речници в DELA формат (Dictionnaires électroniques du LADL – Laboratoire d'Automatique Documentaire et Linguistique), свързани с името на Морис Грос. Формализмът включва създаването на речниците в DELAS и DELAC формат (съответно за прости, сложни и съставни лексикални единици) и дефинирането на морфологични и морфосинтактични локални граматика, които позволяват генерирането на DELAF и DELAFC речници, където са изброени всички форми на простите, сложните и съставните думи с прилежащите им леми и съответната граматична информация. Записът на думите в DELAS и DELAC речниците показва коя дума (основна форма) с коя локална граматика се съотнася и съответно коя локална граматика се използва за разпознаване на дадена форма или за генериране на всички форми на дадена дума. Например записът¹⁰:

мълчà катò рѝба, VPP+I7_P-NCF+15

спя катò пѝн, VPI+I8_P-NCM+16

означава, че глаголните фразеологизми *мълча като риба* и *спя като пън* се асоциират с две различни локални граматика, чиито имена отговарят на формалните записи VPII7_P-NCF+15 и VPII8_P-NCM+16.

Автоматично генерираните записи в DELAF и DELAFC речниците представляват всички форми към дадена основна форма от речниците DELAS и DELAC. Отделният запис се състои от формата на думата, нейната лема (на второ място), обозначения за граматичните категории на лемата (след точката), които предопределят парадигмата на думата, и обозначения за стойностите на граматичните категории на формата (след двете точки). Например:

мълча̀ като̀ риба̀, мълча̀ като̀ риба̀.VPII7_P-NCF+15:1sp_|-fs0
 мълчиш̀ като̀ риба̀, мълча̀ като̀ риба̀.VPII7_P-NCF+15:2sp_|-fs0
 спиш̀ като̀ пън̀, спя̀ като̀ пън̀. VPII8_P-NCM+16:2sp_|-ms0.

Обозначенията за граматичните категории са конвенции, приети в рамките на конкретното приложение. В случая „V“ означава глагол, „P“ на втора позиция – личен, „I“ на трета позиция – свършен вид, „I“ на четвърта позиция – непреходен, „R“ – предлог, „N“ – съществително, „C“ на втора позиция – нарицателно, „F“ на трета позиция – женски род, „M“ на трета позиция – мъжки род, „f“ – женски род, „m“ – мъжки род, „s“ – единствено число, „0“ – неопределено, „1“ – първо лице, „2“ – второ лице, „p“ на трета позиция – сегашно време. Главните букви се използват в тази конвенция за обозначаване на стойностите на граматичните категории на лемата, а малките букви и цифрите – за стойностите на граматичните категории на формата. Символите „_“ и „-“ отбелязват различни видове празни позиции, а „|“ – че думата е неизменяема. Таговете са позиционни, едно и също обозначение в различна позиция може да е стойност на различна категория. Цифрата след знака „плюс“ обозначава локалната граматика за компонент на съставната лексикална единица, цифрата след главна буква означава флективния тип на компонента.

Представянето на локалните граматика може да стане по различен начин: в текстов файл с различна конвенция за обозначаване на изтриването и вмъкването на букви или под формата на граф, който съответства на краен преобразувател. Двете представяния са взаимозаменяеми и могат да се преобразуват едно в друго.

Стойностите на граматичните характеристики на лемата определят възможните форми в рамките на дадена категория (част на речта) и реализираните форми (в рамките на даден подклас, например несвършен и свършен вид) и са част от имената на локалните граматика. Стойностите на граматичните характеристики на формата се приписват от крайния преобразувател при разпознаването или генерирането ѝ и са част от локалната граматика.

5. Категориална и парадигматична информация

Категориалната и парадигматичната информация, част от имената на локалните граматика, определят съответно групирането на единиците в гра-

матични класове (т.е. съществителни, глаголи и т.н.) и в граматични подкласове (т.е. съществителни нарицателни, лични преходни глаголи от свършен вид и т.н.). Парадигматичната информация на съставните лексикални единици включва: броя на компонентите; техния основен словоред (с изключение на конвенцията за кратките притежателни форми); частта на речта (категориалната информация); стойностите на граматичните категории на основната форма на компонентите, които реализират пълната си парадигма; стойностите на граматичните категории на най-неутралната фразеологизирана форма на компонентите, които участват с ограничена парадигма. В парадигматичната информация се включва и дефинирането на празните позиции, които също се приемат за компоненти на фразеологизмите, тъй като означават границите между конституентите на фразеологизмите: думи или словосъчетания.

Поради природата си на граница между конституентите на съставните лексикални единици празните позиции могат да бъдат натоварени с информация дали конституентите, които свързват, могат да разменят местата си, или не, т.е. дали свързват съставни части с фиксиран словоред, или не (Коева 2006: 214), и по този начин да се зададат възможностите за синтактично словоредно вариране. Наблюденията показват, че възможностите за словоредни размествания са пряко свързани с възможностите за вмъкване на елемент в дадена празна позиция.

Затова, за да се представят допустимият словоред и допустимото вмъкване на външни елементи, е достатъчно в рамките на парадигматичната информация да се опишат празните позиции (Коева/Коева 2024: 67): **фиксирана празна позиция** (означава се с „-“), в която не се допуска вмъкване на елемент и словоредна размяна; **фиксирана клитична позиция** (означава се с „|“), в която се позволява вмъкване на притежателна клитика само след членувани форми, но не се допуска словоредна размяна; **фиксирана празна клитична позиция** (означава се с „/“), в която се позволява вмъкване на елемент, но не се допуска словоредна размяна; **празна клитична позиция** (означава се с „=“), в която не се допуска вмъкване на елемент, но се позволява словоредна размяна с клитиката; **празна позиция** (означава се с „_“), в която се позволява вмъкване на елемент и словоредна размяна.

Фиксираната клитична позиция се отнася до именни фразеологизми (*ко-рава глава, коравата му глава*). Фиксираната празна клитична позиция се използва рядко, например за случаи като *не си зная годините* и *не си ли зная годините*. Празната клитична позиция се отнася до притежателната клитика в глаголни фразеологизми и при словоред на клитиката след съществителното в основната форма означава, че клитиката може да заеме позиция преди или след глагола: *тикам под носа му, тикам му под носа, под носа му тикам, нещо му тикам под носа*. Словоредната размяна при празна позиция трябва да се разбира като възможност за размяна на местата на редиците от конституенти, които се свързват до друга празна позиция или край

на фразеологизма. Ако има две **празни позиции** и три редици, това означава, че има шест възможни словореда. Например в глаголният фразеологизъм *тиквам под носа му* има четири лексикални компонента и три вида празни позиции. Между *тиквам* и *под носа му* се кодира **празна позиция**, което означава, че се допускат случаите: *тиквам бързо под носа му* и *под носа му тиквам*. Между *под* и *носа* имаме **фиксирана празна позиция**, което означава, че не се допуска вмъкване на елементи и словоредна размяна. Между *носа* и *му* се кодира **празна клитична позиция**. В глаголният фразеологизъм *вземам хляба от устата му* има пет лексикални елемента и четири празни позиции. Между *вземам* и *хляба* и между *хляба* и *от устата му* има **празни позиции**, което позволява вариантите: *вземам от устата му хляба*, *хляба вземам от устата му*, *от устата му вземам хляба*, *хляба от устата му вземам*, *от устата му вземам хляба*. Между *от* и *устата* има **фиксирана празна позиция**, а между *устата* и *му* – **празна клитична позиция**.

Информацията за вида на празните позиции остава на равнище име на локалната граматика, тъй като, както беше посочено, словоредните промени са регулярни и еднакви за свободните и за устойчивите словосъчетания, ако последните допускат словоредни промени.

При глаголите най-богата е парадигмата от синтетични глаголни форми на личните преходни глаголи от несвършен вид, тъй като от тях се образуват повелителни форми, сегашни деятелни причастия, минали свършени страдателни причастия и деепричастия. Граматичните подкласове при личните глаголи с пълна парадигма на категорията лице са осем, граматичните подкласове при личните глаголи в трето лице единствено и множествено число – 10 и граматичните подкласове при безличните глаголи – 10 (Коева/Коева 1998). Граматичните класове при третоличните глаголи се разглеждат при фразеологизмите, които имат структура на просто изречение. Беше направена уговорката, че както те, така и постоянно рефлексивните, реципрочните, акузативните, дативните и дативно-рефлексивните глаголи остават извън настоящото изследване. Фактически най-често срещаните подкласове са четири: личен преходен несвършен глагол, личен непреходен не-свършен глагол, личен преходен свършен глагол и личен непреходен свършен глагол.

Например основната форма и името на локалната граматика за глаголният фразеологизъм *тиквам под носа му* ще бъдат означени по следния начин:

тѝквам под носà му, VPIT1_R-NCM=PP+4.

Обособени са четири лексикални компонента и три празни позиции, първата позволява вмъкване на елемент и словоредна промяна, втората не позволява вмъкване на елемент и словоредна промяна, а третата не позволява вмъкване на елемент, но позволява словоредна промяна на притежателната клитика. Лексикалните единици са означени като глагол, личен, от

несвършен вид, преходен – „Т“; предлог; съществително, нарицателно от мъжки род; местоимение – „Р“; притежателно за местоимение – „Р“.

Парадигматичната информация (граматичните характеристики на опората, броят и редът на останалите компоненти и техните граматични характеристики, както и характеристиките на празните позиции) специфицира кои от възможните редове от парадигмата на категорията на опората се реализират. Например граматичният клас глагол задава всички възможни редове за глаголите: лични, третолични, безлични; преходни, непреходни; от свършен и несвършен вид.

Редовете на парадигмата може да бъдат увеличени от стойности на категориите на зависимите фразеологизирани компоненти (ако повече от една форма трябва да се съчетае с форма на опората). Например при фразеологизма *ближа си раната* формите се увеличават по две от употребата на съществителното *рана* в единствено число и множествено число: *ближа си раните*. Поради многото варианти, които са реализирани в езика, броят на граматичните подкласове и на флективните типове при фразеологизмите е много по-голям от броя им при думите.

6. Флективна информация

Грамматичният (флективният) тип дефинира по какъв начин се образуват формите: в един флективен тип участват лексикални единици с еднакъв брой членове с еднакви граматични характеристики и еднакви звукови и акцентни редувания (Коева/Коева 1998: 50). Затова лексикални единици с еднаква парадигматична информация може да имат различно флективно описание. Възприет е начин за описанието на промените, настъпващи при формообразуването в отделните лексикални единици, който се базира на изтриване и добавяне на символи, като се записва броят на изтритите наляво символи и експлицитно добавените. Например първите три реда на локалната граматика, описваща флективния тип, към който се отнася глаголният фразеологизъм *сня като пън*, имат следния вид:

```
<>1sp_|-ms0
<1иш>2sp_|-ms0
<1и>3sp_|-ms0
```

Така се кодира, че има промяна само в окончанието на глагола, съответната граматична информация, която трябва да се припише на формата, както и че останалите два компонента в глаголния фразеологизъм не претърпяват промяна.

7. Заключение

Може да се направи обобщението, че за дефинирането на флективните типове на глаголните фразеологизми е необходимо да се дефинира следната парадигматична информация: граматичната информация за основната форма на конституентите, които функционират с пълна парадигма; грама-

тичната информация за най-неутралната фразеологизирана форма на компонентите, които участват с ограничена парадигма; и видът на празните позиции. Празните позиции, които свързват конституентите (думи или словосъчетания) на фразеологизмите, се разглеждат като част от парадигматичното им описание. Дефинирани са **пет** вида празни позиции, за да се осигури представянето на възможностите за промяна на словореда, вмъкването на външни елементи и особеностите при употребата на притежателни клитики.

Доразвива се тезата, че формализираното представяне на флективните типове на простите (сложните) и съставните лексикални единици може да бъде единно: в рамките на един формализъм и с еднакви конвенции. Това улеснява едновременната и независима автоматична обработка, както и създаването на различни компютърни приложения, които не работят само с единия или с другия вид лексикални единици.

Благодарности

Разработката е част от изпълнението на проекта „Българско-полски лексикални и фразеологични апроксимати“, финансиран по грантова схема между Българската академия на науките и Полската академия на науките.

БЕЛЕЖКИ / NOTES

¹ Използваме термините **идиоматичност** и **семантична некомпозиционалност** като синонимни.

² В случая **устойчиви** се употребява по отношение на изразяването от фразеологизмите значение, макар че **устойчиви** е характеристика и по отношение на компонентите на фразеологизмите.

³ Под **флективен (граматичен) тип** се разбира съвкупността от думи с еднаква парадигма, приемана по-широко не само като еднаквост на членовете на парадигмата, но и като еднаквост при звуковите и акцентните редувания (Коева/Коева 1998: 50).

⁴ В тази разработка не се спираме на класификационните типове и на терминологичното разнообразие, като приемаме двата термина: **глаголни съставни лексикални единици** и **глаголни фразеологизми**, за обобщаващи за всички случаи на устойчиви съчетания с глагол. Термините (глаголен) фразеологизъм и (глаголна) съставна лексикална единица се употребяват като синоними в статията (за терминологичното разнообразие и съответствия вж. Калдиева-Захариева/Kaldieva-Zaharieva 2013: 25–29 и др.).

⁵ Представеното описание от В. Малджиева (Малджиева/Maldzhieva 2017) обобщава добре предходни изследвания.

⁶ Обобщено представяне на изброените свойства може да бъде намерено у (Constant et al. 2017). За български характеристиките са описани напълно или частично в редица трудове: Кювлиева-Мишайкова/Kyuvlieva-Mishaykova 1986: 33–34; Ничева/Nicheva 1987: 20–43, 72–77; Калдиева-Захариева/Kaldieva-Zaharieva 2013: 29–32; Коева 2006 и др.

⁷ Структурата на глаголните фразеологизми в български е представена в редица изследвания: Кювлиева-Мишайкова/Kyuvlieva-Mishaykova 1986: 27–28; Maldžieva, Wójtowicz 1994: 66–80; Калдиева-Захаријева/Kaldieva-Zaharieva 2013: 144–167; Тодорова 2015 и др.

⁸ Приема се, че абстрактна лема на фразеологизма представлява последователност от конституенти, които не са леми сами по себе си (Тодорова/Todorova 2015: 94). По-нататък се отбелязва, че всеки конституент на фразеологизма се представя с най-ненатоварената с граматични признаци форма, в която при съчетаване с формите на останалите конституенти се запазва фразеологичността (Тодорова/Todorova 2015: 97). В много фразеологизми обаче може да е фразеологизирана именно форма, съпадаща с основната форма на главната част и/или на конституентите.

⁹ За сравнение са представени български и полски глаголни фразеологизми, в които от линейния ред на думите не може да се направи заключение за граматичната характеристика на съпадащия конституент, а отстраняването на многозначността може да се отбележи при представяне на йерархичната структура:

*Od dawna stara się otworzyć **nam** oczy na wszelkie niesprawiedliwości, jakie dzieją się w tym miejscu pracy* и *Той отдавна се опитва да **ни** отвори очите за всички несправедливости, които се случват на това работно място.*

*Kiedy pokazał **nam**, że na tym dokumencie znajduje się jego podpis, oczy **nam** wyszły z orbit* и *Когато ни показа, че неговият подпис е на този документ, очите **ни** излязоха от орбитите си.*

¹⁰ Ударението е отделен символ пред буквата, която е под ударение.

ЛИТЕРАТУРА

- Благодеева и др. 2022: *Благодеева, Д., С. Коева, В. Сосоновски, М. Яскот*. Речник на активната фразеология в българския и полския език. София, Издателство на Българската академия на науките „Проф. Марин Дринов“.
- Бъркалова 1997: *Бъркалова, П.* Българският синтаксис: познат и непознат. Пловдив, Университетско издателство „Паисий Хилендарски“.
- Виноградов 1947: *Виноградов, В. В.* Русский язык. Грамматическое учение о слове. Ленинград, Государственное учебно-педагогическое издательство.
- Калдиева-Захаријева 2013: *Калдиева-Захаријева, С.* Българска фразеология. София, Академично издателство „Проф. Марин Дринов“.
- Коева 1998: *Коева, С.* Граматичен речник на българския език. Описание на концепцията за организацията на лингвистичните данни. – *Български език*, № 6, с. 49–58.
- Коева 1999: *Коева, С.* Синтаксис и пунктуация. Пловдив, Макрос-2000.
- Коева 2008: *Коева, С.* Българският ФреймНет. Семантико-синтактичен речник на българския език – концептуален модел. – В: Коева, С. (съст.). *Българският ФреймНет: Семантико-синтактичен речник на българския език*. София, Институт за български език „Проф. Любомир Андрейчин“, с. 5–57.
- Коева 2024: *Коева, С.* Речник на активната фразеология в българския и полския език: особености на флективните типове на българските лексикални единици. – В: Колковска, С. (съст. и ред.). *Фразеологични студии*. София, Издателство на Българската академия на науките „Проф. Марин Дринов“, с. 60–78.

- Куцаров 2007: *Куцаров, И.* Теоретична граматика на българския език. Морфология. Пловдив, Коала прес.
- Кювлиева-Мишайкова 1986: *Кювлиева-Мишайкова, В.* Фразеологизмите в българския език. София, Народна просвета.
- Малджиева 2017: *Малджиева, В.* Принципи на формалното описание на полисегментните езикови единици в Нормативната граматика на българския език. – *Български език*, № 2, с. 7–19.
- Ничева 1987: *Ничева, К.* Българска фразеология. София, Наука и изкуство.
- Ничева и др. 1974: *Ничева, К., С. Спасова-Михайлова, К. Чолакова.* Фразеологичен речник на българския език. Т. 1. София, Издателство на Българската академия на науките.
- Паскалева 2007: *Паскалева, Е.* Компютърна морфология. Ресурси и инструменти. София, Институт за паралелна обработка на информацията.
- Пенчев 1998: *Пенчев, Й.* Синтаксис на съвременния български книжовен език. Асеновград, Издателска къща „Вечерник“.
- Стоянова 2012: *Стоянова, И.* Автоматично разпознаване и тагиране на съставни лексикални единици в българския език. Докторска дисертация. София, Институт за български език „Проф. Любомир Андрейчин“.
- Стоянова, Тодорова 2014: *Стоянова, И., М. Тодорова.* Разработване на речници на съставните лексикални единици в българския език за целите на компютърната лингвистика. – В: Коева, С. (съст.). *Езикови ресурси и технологии за български език*. София, Академично издателство „Проф. Марин Дринов“, с. 185–202.
- Тодорова 2015: *Тодорова, М.* Типология и свойства на устойчиви словосъчетания в български. Глаголни фразеологизми. Докторска дисертация. София, Институт за български език „Проф. Любомир Андрейчин“.
- Baldwin, Kim 2010: *Baldwin, T., S. N. Kim.* Multiword expressions. – In: Indurkha, N., F. J. Damerau (eds.). *Handbook of Natural Language Processing*. 2nd ed. Boca Raton, CRC Press, pp. 267–292.
- Bally 1909: *Bally, Ch.* *Traité de stylistique française*. Vol. 1. Paris, C. Winter.
- Benveniste 1974: *Benveniste, É.* Formes nouvelles de la composition nominale. – In: Benveniste, É. (ed.). *Problèmes de linguistique générale*, vol. 2. Paris, Gallimard, pp. 163–176.
- Breidt et al. 1996: *Breidt, E., F. Segond, G. Valetto.* Formal description of multi-word lexemes with the finite-state formalism IDAREX. – In: *16th International Conference on Computational Linguistics. Proceedings of the Conference, COLING 1996*. Copenhagen, Center for Sprogteknologi, pp. 1036–1040.
- Chrobot et al. 1999: *Chrobot, A., B. Courtois, M. Herman, M. Gross, K. Zellagui.* Dictionnaire Electronique DELAC anglais: noms composés. Rapport technique n°59, LADL, Université Paris 7.
- Constant et al. 2017: *Constant, M., G. Eryigit, J. Monti, L. van der Plas, C. Ramisch, M. Rosner, A. Todirascu.* Multiword expression processing: A survey. – *Computational Linguistics*, 43 (4), pp. 837–892.

- Courtois, Silberztein 1990: *Courtois, B., M. Silberztein*. Dictionnaires électroniques du français. – *Langue française*, 87 (1), pp. 11–22.
- Fellbaum 1993: *Fellbaum, C.* The determiner in English idioms. – In: Cacciari, C., P. Tabossi (eds.). *Idioms: Processing, Structure and Interpretation*. Hillsdale, NJ, Lawrence Erlbaum Associates, pp. 271–296.
- Grégoire 2010: *Grégoire, N.* DuELME: a Dutch electronic lexicon of multiword expressions. – *Language Resources and Evaluation*, 44, pp. 23–39.
- Gross 1990: *Gross, G.* Définition des noms composés dans un lexique-grammaire. – *Langue Française*, 87, pp. 84–90.
- Gross 1997: *Gross, M.* The construction of local grammars. – In: Roche E., Y. Schabès (eds.). *Finite-State Language Processing*. MIT Press, pp. 329–354.
- Koeva 2006: *Koeva, S.* Inflection morphology of Bulgarian multiword expressions. – In: *Computer Applications in Slavic Studies*. Sofia, Boyan Penev Publishing House, pp. 201–216.
- Koeva et al. 2016: *Koeva, S., I. Stoyanova, M. Todorova, S. Leseva*. Semi-automatic compilation of the Dictionary of Bulgarian Multiword Expressions. – In: *Proceedings of the GLOBALEX 2016 Workshop: Lexicographic Resources for Human Language Technology*, LREC, pp. 86–95.
- Lion-Bouton et al. 2023: *Lion-Bouton, A., A. Savary, J.-Y. Antoine*. A MWE lexicon formalism optimised for observational adequacy. – In: *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*. Dubrovnik, Association for Computational Linguistics, pp. 121–130.
- Maldżieva, Wójtowicz 1994: *Maldżieva, V., J. Wójtowicz*. Polskie i bułgarskie związki frazeologiczne. Warszawa, Wydawnictwo Uniwersytetu Warszawskiego.
- Nunberg et al. 1994: *Nunberg, G., I. Sag, T. Wasow*. Idioms. – *Language*, 70, pp. 491–538.
- O’Grady 1998: *O’Grady, W.* The syntax of idioms. – *Natural Language and Linguistic Theory*, 16, pp. 279–312.
- Oflazer et al. 2004: *Oflazer, K., Ö. Çetinoğlu, B. Say*. Integrating morphology with multi-word expression processing in Turkish. – In: *Proceedings of the Workshop on Multiword Expressions: Integrating Processing*. Barcelona, Spain, Association for Computational Linguistics, pp. 64–71.
- Przepiórkowski et al. 2017: *Przepiórkowski, A., J. Hajič, E. Hajnicz, Z. Urešová*. Phraseology in two Slavic valency dictionaries: Limitations and perspectives. – *International Journal of Lexicography*, 30 (1), pp. 1–38.
- Sag et al. 2002: *Sag, I. A., T. Baldwin, F. Bond, A. Copestake, D. Flickinger*. Multiword expressions: A pain in the neck for NLP. – In: Gelbukh, A. (ed.). *Computational Linguistics and Intelligent Text Processing. CICLing 2002. Lecture Notes in Computer Science*. Vol. 2276, Berlin/Heidelberg, Springer, pp. 189–206.
- Sailer, Trawiński 2006: *Sailer M., B. Trawiński*. The collection of distributionally idiosyncratic items: A multilingual resource for linguistic research. – In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06), Genoa, Italy*. European Language Resources Association (ELRA), pp. 471–474.

- Savary 2005: *Savary, A.* A formalism for the computational morphology of multi-word units. – *Archives of Control Sciences*, 15(3), pp. 437–449.
- Savary 2008: *Savary, A.* Computational inflection of multi-word units. – In: *Linguistic Issues in Language Technology*, CSLI Publications, 1 (2), pp. 1–53.
- Savary et al. 2018: *Savary, A., S. Petitjean, T. Lichte, L. Kallmeyer, J. Waszczuk.* Object-oriented lexical encoding of multiword expressions: Short and sweet. arXiv preprint arXiv:1810.09947.
- Silberztein 1993: *Silberztein, M.* Les groupes nominaux productifs et les noms composés lexicalisés. – *Linguisticae Investigationes*, 17(2), pp. 405–425.
- Silberztein 2005: *Silberztein, M.* NooJ's dictionaries. – In: *Proceedings of Language and Technology Conference, Poznań*. Poznań, Wydawnictwo Poznańskie, pp. 291–295.
- Vondříčka 2019: *Vondříčka, P.* Design of a multiword expressions database. – In: *The Prague Bulletin of Mathematical Linguistics*, 112, pp. 83–101.
- Zgusta 1967: *Zgusta, L.* Multiword lexical units. – *Word*, 23 (1–3), pp. 578–587.

REFERENCES

- Baldwin, Kim 2010: *Baldwin, T., S. N. Kim.* Multiword expressions. – In: Indurkha, N., F., J. Damerau (eds.). *Handbook of Natural Language Processing*, 2nd ed. Boca Raton, CRC Press, pp. 267–292.
- Bally 1909: *Bally, Ch.* Traité de stylistique française. Volume 1. Paris, C. Winter.
- Barkalova 1997: *Barkalova, P.* Balgarskiyat sintaksis: poznat i nepoznat. Plovdiv, Universitetsko izdatelstvo “Paisiy Hilendarski”.
- Benveniste 1974: *Benveniste, É.* Formes nouvelles de la composition nominale. – In: Benveniste, É. (ed.). *Problèmes de linguistique générale*. Vol. 2. Paris, Gallimard, pp. 163–176.
- Blagoeva et al. 2022: *Blagoeva, D., S. Koeva, V. Sosonovski, M. Yaskot.* Rechnik na aktivnata frazeologiya v balgarskiya i polskiya ezik, Sofia, Izdatelstvo na Balgarskata akademiya na naukite “Prof. Marin Drinov”.
- Breidt et al. 1996: *Breidt, E., F. Segond, G. Valetto.* Formal description of multi-word lexemes with the finite-state formalism IDAREX. – In: *16th International Conference on Computational Linguistics. Proceedings of the Conference, COLING 1996*. Copenhagen, Center for Sprogteknologi, pp. 1036–1040.
- Chrobot et al. 1999: *Chrobot, A., B. Courtois, M. Herman, M. Gross, K. Zellagui.* Dictionnaire Electronique DELAC anglais: noms composés. Rapport technique n°59, LADL, Université Paris 7.
- Constant et al. 2017: *Constant, M., G. Eryiğit, J. Monti, L. van der Plas, C. Ramisch, M. Rosner, A. Todirascu.* Multiword expression processing: A survey. – *Computational Linguistics* 43:4, pp. 837–892.
- Courtois, Silberztein 1990: *Courtois, B., M. Silberztein.* Dictionnaires électroniques du français. – *Langue française*, 87 (1), pp. 11–22.
- Fellbaum 1993: *Fellbaum, C.* The determiner in English idioms. – In: Cacciari, C., P. Tabossi (eds.). *Idioms: Processing, structure and interpretation*. Hillsdale, NJ, Lawrence Erlbaum Associates, pp. 271–296.

- Grégoire 2010: *Grégoire, N.* DuELME: a Dutch electronic lexicon of multiword expressions. – *Language Resources and Evaluation*, 44, pp. 23–39.
- Gross 1990: *Gross, G.* Définition des noms composés dans un lexique-grammaire. – *Langue Française*, 87, pp. 84–90.
- Gross 1997: *Gross, M.* The construction of local grammars. – In: Roche E., Y. Schabès (eds.). *Finite-State Language Processing*. MIT Press, pp. 329–354.
- Kaldieva-Zaharieva 2013: *Kaldieva-Zaharieva, S.* Балгарска фразеологија. София, Академично издателство “Prof. Marin Drinov”.
- Koeva 1998: *Koeva, S.* Граматичен реchnik на балгарскија език. Описание на контсептијата за организацијата на лингвистичните данни. – *Balgarski ezik*, N 6, s. 49–58.
- Koeva 1999: *Koeva, S.* Синтаксис и пунктуација. Пловдив, Makros-2000.
- Koeva 2006: *Koeva, S.* Inflection morphology of Bulgarian multiword expressions. – In: *Computer Applications in Slavic Studies*. София, Boyan Penev Publishing House, pp. 201–216.
- Koeva 2008: *Koeva, S.* Балгарският FrameNet. Семантико-синтактичен реchnik на балгарскија език – контсептуален модел. – In: *Koeva, S. (sast.). Balgarskiyat FrameNet: Semantiko-sintaktichen rechnik na balgarskiya ezik*. София, Institut za balgarski ezik “Prof. Lyubomir Andreychin”, s. 5–57.
- Koeva 2024: *Koeva, S.* Реchnik на активната фразеологија во балгарскија и полскија език: особености на флективните типови на балгарските лексикални единици. – In: *Kolkovska, S. (sast. i red.). Frazeologichni studii*. София, Izdatelstvo na Balgarskata akademiya na naukite “Prof. Marin Drinov”, s. 60–78.
- Koeva et al. 2016: *Koeva, S., I. Stoyanova, M. Todorova, S. Leseva.* Semi-automatic compilation of the Dictionary of Bulgarian Multiword Expressions. – In: *Proceedings of the GLOBALEX 2016 Workshop: Lexicographic Resources for Human Language Technology*, LREC, pp. 86–95.
- Kutsarov 2007: *Kutsarov, I.* Теоретична граматика на балгарскија език. Морфологија. Пловдив, Koala pres.
- Kyuvlieva-Mishaykova 1986: *Kyuvlieva-Mishaykova, V.* Фразеологизмите во балгарскија език. София, Narodna prosveta.
- Lion-Bouton et al. 2023: *Lion-Bouton, A., A. Savary, J.-Y. Antoine.* A MWE lexicon formalism optimised for observational adequacy. – In: *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*. Dubrovnik, Association for Computational Linguistics, pp. 121–130.
- Maldzhieva 2017: *Maldzhieva, V.* Принципи на формалното описание на полисеметните езикови единици во Нормативната граматика на балгарскија език. – *Balgarski ezik*, N 2, s. 7–19.
- Maldzhieva, Wójtowicz 1994: *Maldzhieva, V., J. Wójtowicz.* Польские и българские звязки фразеологичне. Warszawa, Wydawnictwo Uniwersytetu Warszawskiego.
- Nicheva 1987: *Nicheva, K.* Балгарска фразеологија. София, Nauka i izkustvo.
- Nicheva et al. 1974: *Nicheva, K., S. Spasova-Mihaylova, K. Cholakova.* Фразеологичен реchnik на балгарскија език. Т. 1. София, Izdatelstvo na Balgarskata akademiya na naukite.
- Nunberg et al. 1994: *Nunberg, G., I. Sag, T. Wasow.* Idioms. – *Language*, 70, pp. 491–538.

- O'Grady 1998: *O'Grady, W.* The syntax of idioms. – *Natural Language and Linguistic Theory*, 16, pp. 279–312.
- Oflazer et al. 2004: *Oflazer, K., Ö. Çetinoğlu, B. Say.* Integrating morphology with multi-word expression processing in Turkish. – In: *Proceedings of the Workshop on Multiword Expressions: Integrating Processing*. Barcelona, Spain, Association for Computational Linguistics, pp. 64–71.
- Paskaleva 2007: *Paskaleva, E.* Kompyutarna morfologiya. Resursi i instrumenti. Sofia, Institut za paralelna obrabotka na informatsiyata.
- Penchev 1998: *Penchev, Y.* Sintaksis na savremenniya balgarski knizhoven ezik. Ase novgrad, Izdatelska kashta “Vechernik”.
- Przepiórkowski et al. 2017: *Przepiórkowski, A., J. Hajič, E. Hajnicz, Z. Urešová.* Phraseology in two Slavic valency dictionaries: Limitations and perspectives. – *International Journal of Lexicography*, 30 (1), pp. 1–38.
- Sag et al. 2002: *Sag, I. A., T. Baldwin, F. Bond, A. Copestake, D. Flickinger.* Multiword expressions: A pain in the neck for NLP. – In: Gelbukh, A. (ed.). *Computational Linguistics and Intelligent Text Processing. CICLing 2002. Lecture Notes in Computer Science*. Vol. 2276, Berlin/Heidelberg, Springer, pp. 189–206.
- Sailer, Trawiński 2006: *Sailer M., B. Trawiński.* The collection of distributionally idiosyncratic items: A multilingual resource for linguistic research. – In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy*. European Language Resources Association (ELRA), pp. 471–474.
- Savary 2005: *Savary, A.* A formalism for the computational morphology of multiword units. – *Archives of Control Sciences*, 15(3), pp. 437–449.
- Savary 2008: *Savary, A.* Computational inflection of multi-word units. – In: *Linguistic Issues in Language Technology*, CSLI Publications, Vol. 1, Issue 2, pp. 1–53.
- Savary et al. 2018: *Savary, A., S. Petitjean, T. Lichte, L. Kallmeyer, J. Waszczuk.* Object-oriented lexical encoding of multiword expressions: Short and sweet. arXiv preprint arXiv:1810.09947.
- Silberstein 1993: *Silberstein, M.* Les groupes nominaux productifs et les noms composés lexicalisés. – *Linguisticae Investigationes*, 17(2), pp. 405–425.
- Silberstein 2005: *Silberstein, M.* NooJ's dictionaries. – In: *Proceedings of Language and Technology Conference, Poznań*. Poznań, Wydawnictwo Poznańskie, pp. 291–295.
- Stoyanova 2012: *Stoyanova, I.* Avtomatichno razpoznavane i tagirane na sastavni leksikalni edinitsi v balgarskiya ezik. Doktorska disertatsiya. Sofia, Institut za balgarski ezik “Prof. Lyubomir Andreychin”.
- Stoyanova, Todorova 2014: *Stoyanova, I., M. Todorova.* Razrabotvane na rechnitsi na sastavnite leksikalni edinitsi v balgarskiya ezik za tselite na kompyutarnata lingvistika. – In: Koeva, S. (sast.). *Ezikovi resursi i tehnologii za balgarski ezik*. Sofia, Akademichno izdatelstvo “Prof. Marin Drinov”, s. 185–202,
- Todorova 2015: *Todorova, M.* Tipologiya i svoystva na ustoychivi slovosachetaniya v balgarski. Glagolni frazeologizmi. Doktorska disertatsiya. Sofia, Institut za balgarski ezik “Prof. Lyubomir Andreychin”.

Vinogradov 1947: *Vinogradov, V. V.* Russkii iazyk. Grammaticheskoe uchenie o slove. Leningrad, Gosudarstvennoe uchebno-pedagogicheskoe izdatel'stvo.

Vondřička 2019: *Vondřička, P.* Design of a multiword expressions database. – In: *The Prague Bulletin of Mathematical Linguistics*, 112, pp. 83–101.

Zgusta 1967: *Zgusta, L.* Multiword lexical units. – *Word*, 23 (1–3), pp. 578–587.

✉ *Проф. д-р Светла Коева*

Секция по компютърна лингвистика

Институт за български език „Проф. Любомир Андрейчин“

при Българската академия на науките

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

✉ *Prof. Dr. Svetla Koeva*

Department of Computational Linguistics

Institute for Bulgarian Language, Bulgarian Academy of Sciences

52 Shipchenski prohod, Bl. 17, 1113 Sofia, Bulgaria