Български език, Приложение / Bulgarian Language, Supplement, 72 (2025), 120–132
Print ISSN: 0005-4283, Online ISSN: 2603-3372
doi: 10.47810/BL.72.25.PR.09

Флективно описание на съставни наименования в българския език (изследване на двусъставните географски названия)

Светлозара Лесева
Институт за български език „Проф. Любомир Андрейчин“ при Българската академия на науките
zarka@dcl.bas.bg

В статията се разглеждат граматичните и семантичните особености на двусъставните наименования в българския език с оглед на класифицирането им към един или друг флективен тип, като фокусът е върху наименования на географски обекти. Единиците, върху които са извършени наблюденията, са извлечени автоматично от интернет, включително от структурирани ресурси като Уикипедия, след което са филтрирани ръчно и им е приписана категория според типа на референта, например *град*, *село*, *планина*, *извор* и др. В резултат от филтрирането броят на наименованията е редуциран от 4000 на близо 2600. Единиците са тагирани автоматично, в резултат от което на всеки компонент на всяко наименование са приписани етикети за част на речта и морфологични характеристики.

Наименованията се разглеждат комплексно в няколко аспекта, свързани със синтактичната им структура, лексикално-граматичните категории и формоизменителните им особености и т.н. В своята съвкупност тези характеристики предопределят флективния тип на наименованията.

Представеното описание служи както за генериране на словоформите на съставните наименования, така и за разпознаване на формите им и приписване на релевантна граматична информация в свободен текст.

Ключови думи: флективен тип; географски названия; съставни наименования; български език

Inflectional Description of Bulgarian Multiword Expressions: An Analysis of Two-Component Geographical Names

Svetlozara Leseva
Institute for Bulgarian Language, Bulgarian Academy of Sciences
zarka@dcl.bas.bg

This article explores the grammatical and semantic characteristics of two-component compound proper names in Bulgarian, with a particular focus on geographical names, aiming to define their inflectional types and integrate them into the respective dictionary

entries. The study is based on lexical units automatically extracted from internet sources, including structured resources such as Wikipedia. These units were manually filtered and categorised according to the type of referent, e.g. *town*, *village*, *mountain*, *spring*, etc. As a result, the initial dataset of 4,000 candidates was reduced to approximately 2,600 entries.

The selected names were automatically POS-tagged and analysed in terms of their syntactic structure, lexical and grammatical categories, inflectional properties, defectivity, etc. These features collectively determine the inflectional type of each named entity.

The inflectional types enable both the generation of inflected forms and their recognition in free text.

Keywords: inflection types, geographical names, multiword expressions, Bulgarian

1. Въведение

В статията се разглеждат граматичните и семантичните особености на двусъставните наименования в българския език с оглед на класифицирането им към един или друг флективен тип, като фокусът е върху наименования на географски обекти. Единиците, върху които са извършени наблюденията, са извлечени автоматично от интернет, включително от структурирани ресурси като Уикипедия, след което са филтрирани ръчно и им е приписана категория според типа на референта (например *град*, *село*, *планина*, *извор* и др.). В резултат на това броят на наименованията е редуциран от 4000 на близо 2600 единици. Наименованията са тагирани автоматично, в резултат от което на всеки компонент са приписани част на речта и морфологични характеристики.

На следващия етап от работата наименованията са разгледани комплексно в няколко аспекта, предопределящи флективния тип, а именно вътрешната им структура, лексикално-граматичните разреди, към които се отнасят компонентите им, лексикално-граматичните категории на компонентите им, формоизменителните им особености. В съвкупността си тези аспекти предопределят подялбата на съставните наименования във флективни типове.

В изследването са включени както изменяеми, така и неизменяеми наименования от домашен и от чужд произход. Интересно предизвикателство представлява описанието на наименования от чужд произход, към които се отнасят както такива, чиято вътрешна структура е ясна и сводима към вече идентифицирани типове (например *Свети Иван* и *Санкт Петербург*), така и други, при които се изисква експертно знание за конкретните езици и контекст на употреба.

Описанието служи както за генериране на словоформите на съставните наименования, така и за разпознаване на формите им и приписване на релевантна граматична информация в свободен текст.

2. Актуалност на проблематиката

Съставните лексикални единици, към които се отнасят и съставните наименования, представляват едно от основните частни проблемни полета в областта на компютърната лингвистика и свързаните с нея приложения и

езикови технологии. Главните предизвикателства, които това езиково явление поставя, са свързани със значителната степен на *семантична непрозрачност*, т.е неизводимост на значението от значението на компонентите при част от тях; *лексикална невариативност,* т.е. невъзможност за замяна на някой от компонентите с негов синоним, въпреки че съставното значение е семантично прозрачно и изводимо от значението на компонентите; *морфологична или синтактична непредсказуемост* на форми на компонентите или на цялата единица; *пропускливост/непропускливост* за външни за съставната лексикална единица елементи; съгласувателни особености и др. (Sag et al. 2002).

Предизвикателствата, произтичащи от езиковите свойства на съставните лексикални единици както в теоретичен план, така и от гледна точка на откриването и разпознаването им в текст от системите за обработка на естествения език, все още не са намерили окончателно решение. Разработването на богати бази данни от съставни лексикални единици продължава да бъде основна предпоставка за коректното генериране на формите от дадена основна форма и достатъчно доброто разпознаване на съставните единици в свободен текст (Savary et al. 2019). През последните години този научен проблем получава нови измерения, а именно – усъвършенстването и адаптацията на големите езикови модели чрез интегрирането на лексикографска информация (вж. Dimakis et al. 2024).

Въпросът за пълното и последователно описание на съставните единици от гледна точка на генерирането на формите от парадигмата им има солидна традиция и с оглед на българския език (вж. напр. Koeva et al. 2016). Неговата актуалност се предопределя и от съвременните тенденции в лексикографията и кодификацията на създаваните онлайн ресурси, където стремежът е да се включва все по-изчерпателно описание с информация за правописа, произношението и граматичните характеристики на лексемите под формата на разгърнати парадигми. Такива ресурси са *Граматичният речник на полския език*[1], *Граматичният речник на руския език*[2], *Граматичният речник на българския език* (Коева/Koeva 1998), *Официалният правописен речник на българския език* (*онлайн версия)*[3] и др.

По-долу ще се спрем на част от спецификите на съставните единици, които дават отражение върху описанието им в граматичен речник.

3. Езикови особености на съставните наименования

Съставните наименования могат да бъдат както неизменяеми, така и изменяеми, като тази специфика не се определя единствено от това дали в състава на наименованието има собствено име (*Долна Митрополия, Долно Градище*), или не (*Долна баня*). Така например, подобно на личните имена, имената на населените места (градове, села, махали) са неизменяеми независимо от състава си. От друга страна, имената на много от географските обекти като равнини, низини, планини запазват свойствата на синтактич-

ната си структура и най-често могат да се членуват. Това обаче също не е безизключително: например наименованията на моретата в общия случай остават неизменяеми (*Яванско море*), а членуваната форма, доколкото е синтактично възможна, е оказионална и стилистично маркирана.

Тук трябва да се вземе предвид и фактът, че едни и същи имена може да се отнасят към различни референти, при това е възможно те да са засвидетелствани в практиката и кодирани в речниците в различна от неутралната нечленувана форма. Така например се срещат топоними, които са кодифицирани в различни варианти с определителен член, срв. *Синият вир* (Заимов/Zaimov 2021: 81) и *Синия вир* (Заимов/Zaimov 2021:80), *Големият скок* (Заимов/Zaimov 2012: 291) и *Големия скок* (Заимов/Zaimov 2012: 290)[4]. Изясняването на тези особености и отразяването на непредсказуемите, идеосинкретични характеристики на конкретните единици е важна предпоставка за правилното им отнасяне към даден флективен тип.

Поради факта, че имат уникален референт, в общия случай географските наименования са неизменяеми по число, като в зависимост от характера на референта могат да бъдат в единствено (*Бургаско езеро*) или в множествено число (*Мальовишки езера*), когато се назовава група обекти. В някои случаи се наблюдава несъответствие между типа на референта и морфологичната маркираност на наименованието по число (възвишение *Самуиловски височини*).

Не на последно място ще посочим, че в съпоставка с други типове съставни лексеми наименованията се характеризират с по-висока степен на ограниченост, например при тях не се наблюдава или се наблюдава рядко разпокъсване на линейното разположение на компонентите: срв. *киселото* ***ви*** *мляко* спрямо **Черепишкия ви манастир*, както и словоредни вариации.

Въпросът за това каква точно е основната форма (при неизменяемите съставни единици) и кои са членовете на парадигмата, е важен както от лингвистична гледна точка, така и за автоматичното откриване на съставните единици в свободен текст за целите на компютърната лингвистика и автоматичната обработка на естествения език. Срещането на форма, която не принадлежи на парадигмата, може да бъде индикация, че става дума за свободно словосъчетание, например: *Жълтото море от слънчогледи ги мамеше,* но *Жълто море ги мамеше с богатствата си*. Въпреки че обикновено се описват като неизменяеми, някои от тези наименования понякога биват подвеждани под общите правила за изменение, важащи за свободните фрази и членуващите се съставни лексикални единици: *Черното море*, *Жълтото море*, под влияние вероятно на прагматични фактори. Само по себе си това не затруднява носителите на езика, но поставя пред създателите на граматични и други речници въпроса дали такива форми трябва да бъдат включени. Този въпрос, подобно на останалите особености, описани в настоящата част, има отношение към избора на флективен тип.

4. Принципи на флективното описание с оглед на географските наименования

Работата по географските наименования продължава и надгражда предприетото по-рано описание на именните съставни единици (Koeva et al. 2016; Тодорова и др./Todorova et al. 2017 и мн. др.). Изхожда се от вече формулираните 81 флективни типа, използвани за генерирането и разпознаването на двусъставни единици.

Флективните типове на съставните лексикални единици се дефинират въз основа на следните характеристики:

а) структурен тип на съставната единица, който представя вътрешната ѝ синтактична структура, например:

A N (прилагателно + съществително име): *Белите брези*, *Момина клисура*, *Айдемирска низина*;

N N (съществително име + съществително име): *Бачо Киро*, *Генерал Тодоров*, *Берекет могила*;

б) лексикално-граматични разреди, към които се отнасят компонентите на съставните наименования, най-релевантните сред които са разредите при съществителните имена: собствени/нарицателни;

в) семантична характеристика, например лични собствени имена, географски собствени имена и др.;

г) лексикално-граматични категории на компонентите: за конкретната задача определящ е родът на съществителното, което е опора на съставното наименование;

д) формоизменителни особености на компонентите: в конкретния случай те включват преди всичко възможността за членуване при изменяемите имена.

С оглед на конкретната задача, а именно – описанието на географски наименования, семантичната характеристика (вж. подточка (в) по-горе) е предварително приписана. В хода на настоящата работа тя е допълнително детайлизирана чрез добавянето на типа географско наименование, например *село*, *град*, *местност*, *водопад* и т.н. Тази информация е извлечена автоматично и впоследствие проверена. Към момента в речника са включени 95 подтипа наименования на природни и административни обекти.

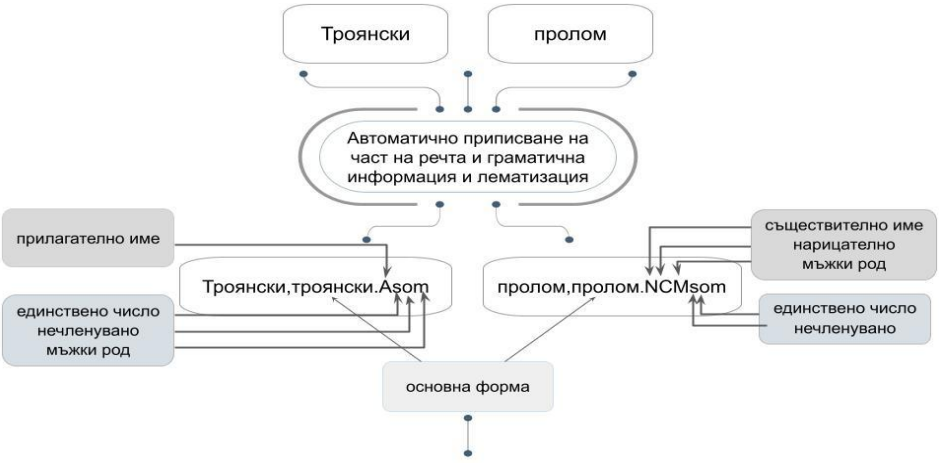
Работата по създаването и приписването на флективен тип включва няколко етапа, които са описани по-долу.

I. Предварителният етап включва филтрирането на единиците, така че да се изключат имената, съдържащи грешно изписване, включително предаване на кирилица в разрез с установените правила за транскрипция на чужди имена. Така например извлеченото от интернет наименование *Айлънд Парк* се среща и с вариантите *Айланд Парк* и *Айлънд парк*, които след проверката са изключени от списъка. Като следваща стъпка е приписан типът географски или административен обект след изрична проверка. Тази

процедура няма пряко отношение към дефинирането и приписването на флективен тип, но от една страна, има висока информационна стойност, а от друга, помага за оразличаване на еднакви единици с различни референти.

Полученият списък с имена е подложен на автоматична обработка с помощта на Многокомпонентната система за лингвистичен анализ за български език (Bulgarian Language Processing Chain) (Koeva et al. 2020). Системата включва и модул, който определя частта на речта и граматичните характеристики на всяка словоформа (тагер), както и лематизатор, който им приписва основна форма (лема). На този етап всяка дума се обработва и се третира самостоятелно, а не като част от съставна лексикална единица. След това анотираните единици се подлагат на полуавтоматична и ръчна проверка, в резултат на което се отстраняват неточно приписаните части на речта и категории.

Този тип проверки и корекция са значителни по обем, тъй като много от компонентите на имената не фигурират в граматичния речник, използван при тагирането и лематизацията, и съответно не могат да бъдат разпознати: диалектни варианти (*Бела Рада*), остатъчни сложни форми на прилагателните (*Жълти рид*), чужди думи и имена (*Гранд Рапидс*, *Домброва Гурнича, Кастелина Маритима*) и др. На този етап се взема решение кои от неразпознатите думи да бъдат въведени в граматичен речник и описани и към кои да се възприеме различна стратегия.



Фигура 1. Приписване на част на речта, граматични характеристики и основна форма на компонентите на съставните единици

На фигура 1 е онагледен резултатът от приписването на част на речта, граматични характеристики и основна форма на всеки от компонентите, като са показани лексикално-граматичните и морфологичните категории, засегнати от формоизменението, и техните стойности.

II. На следващия етап от работата се извършва приписване на единна основна форма на съставната единица, която представлява съчетание от формите на компонентите ѝ във вида, в който реално са засвидетелствани. Процедурата се извършва автоматично, като резултатът се подлага на последваща ръчна проверка. При *Троянски пролом* основната форма на съставната единица съвпада със съчетанието на основните форми на компонентите ѝ. При *Тракийска низина* обаче в основната форма прилагателното е във форма за женски род, а не в мъжки, тъй като, подобно на свободните словосъчетания, прилагателното, изпълняващо функция на определение, се съгласува по род и число с главната дума.

На този етап се приписва и лексикалната и лексикално-граматичната информация, която представлява интегрална част от флективния тип. Релевантни за флективните типове, описващи разглежданите съставни наименования, са лексикално-граматичната категория *род* (в случая м.р., или „M“), която се определя от рода на опорното съществително, и лексикално-граматичният разред, към който принадлежи единицата, а именно – съществително собствено (означено с „H“).

Въпреки че няма пряко отношение към формоизменението и граматичните характеристики на единиците, във флективния тип е включена и тематичната област („G“, или география). Тъй като наименованието има уникален референт и не се изменя по число, тази категория условно може да се разглежда като лексикално-граматична характеристика на съставната лексикална единица („S“).

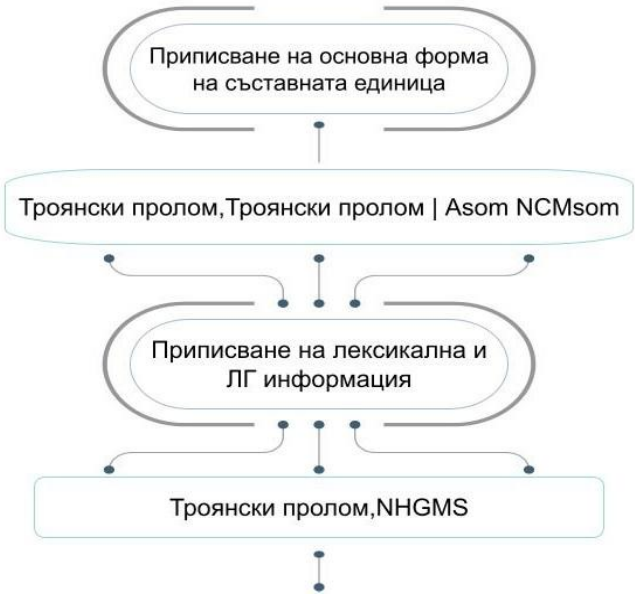
III. При следващия етап от работата се извършва анализ на вътрешната синтактична структура на единицата, както и на възможностите за словоредни вариации и вмъкване на външни елементи между компонентите ѝ и се установяват морфологичните категории, с които е свързано формоизменението. В разглеждания пример структурата включва прилагателно (означено с „A“) и съществително (означено с „N“) нарицателно (означено с „C“) име от мъжки род („M“), между които не може да се вмъкват външни елементи (означава се с подчертаване „_“) и не са позволени словоредни варианти (по подразбиране). Съгласуването между подчинената и главната част на словосъчетанието, отразено в основната форма, от която се генерират останалите, се извършва по признаците род (мъжки – „m**“**), число (единствено – „s“) и определеност (със стойност „o“ за основната форма). На базата на извършения анализ се съставя втората част от наименованието на флективния тип – Asom_MCNsom. Така то (NHGMS:Asom_MCNsom) кодира информация за лексикално-граматичните, морфологичните, синтактичните и морфосинтактичните особености на единицата.

IV. На този етап се анализират особеностите на формоизменението на съответната единица. Въз основа на това се извършва формално описание на елементарните операции, чрез които от основната форма се получават

останалите форми. Това става чрез изменение на формата на прилагателното, като единствената категория, чиито стойности могат да се изменят, е определеността. Тъй като единицата е от мъжки род, то парадигмата ѝ е тричленна и включва основната форма и членуваните форми с пълен и с кратък член.

Генерирането на членуваните форми се извършва чрез конкатенация (слепване, верижно свързване) на основната форма на прилагателното, завършващо на *-ски*, и членната морфема *-ят* или *-я*, без други (морфофонемни) изменения. Тези особености са отразени в допълнителното означение „3-0“, което обозначава поредния формулиран подтип от по-общия тип NHGMS:Asom_MCNsom (фигура 3).

Възприетият граматичен формализъм се основава на добре познати в компютърната лингвистика формати за речниково описание на несъставни (DELAS) и съставни лексеми (DELAC) (Courtois, Silberztein 1990) и на лексикон-граматиките (Gross 1997), адаптирани и разширени за конкретните цели.



Фигура 2. Приписване на основна форма на съставната единица, описание на релевантната лексикална и лексикално-граматична информация и създаване на флективния тип

По-долу е показано и формалното описание на конкретните операции, чрез които се генерира парадигмата не само на разглежданото наименование, но и на всички останали единици, подчиняващи се на същия формообразувателен модел:

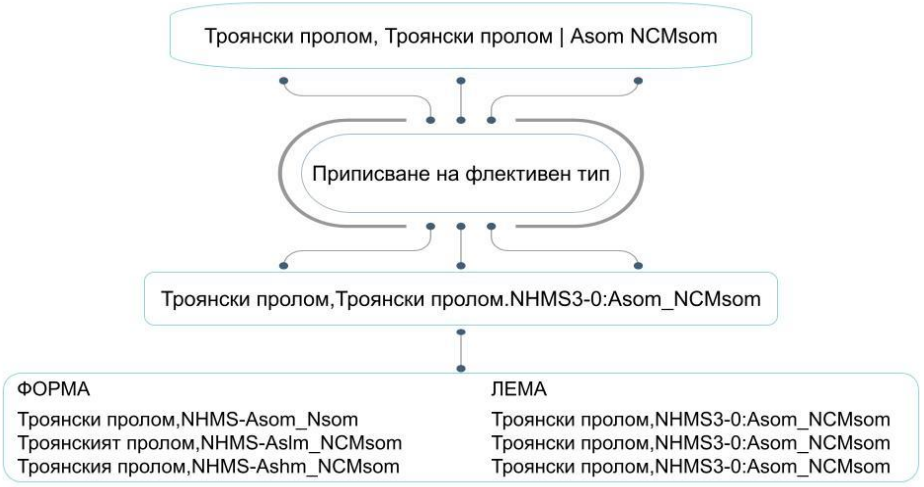
NHMS3-0:Asom_NCMsom =<1> <2>/-Asom_Nsom +
<1>ят <2>/-Aslm_NCMsom +
<1>я <2>/-Ashm_NCMsom

Приложението на тези операции към конкретното наименование изглежда по следния начин (с ∅ е означена нулевата членна морфема, с **l** – членуването с пълен член; с **h** – членуването с непълен член):

Троянски∅ пролом:Asom_Nsom +
Троянски**ят** пролом:Aslm_NCMsom +
Троянски**я** пролом:Ashm_NCMsom

Всеки от редовете във флективното описание кодира двукомпонентната структура: <1><2> и изменението на основната форма чрез членуване на първия компонент (прилагателното име) чрез конкатенация на съответната членна морфема към основната форма на прилагателното.

След наклонената черта към генерираната вече форма на съставната единица и приписания ѝ флективен тип (NHMS3-0) се добавя и граматичното ѝ описание: Aslm_NCMsom: членувано с пълен член прилагателно в единствено число, следвано от нечленувано съществително нарицателно от мъжки род в единствено число, между които не може да се вмъкват външни елементи. В резултат генерираните форми получават разгърнатите означения, посочени на фигура 3.



Фигура 3. Генериране на словоформите по разглеждания формообразувателен модел

Във възприетия формат след генерираната форма и граматичните ѝ характеристики се извежда и лемата заедно с флективния тип (фигура 3).

По аналогичен начин са формулирани и останалите флективни типове и подтипове на съставните наименования, включени в речника.

5. Особености на географските названия във връзка с тяхното флективно описание

Както бе посочено, при географските наименования се наблюдават различни специфики, като основни форми, фиксирани в определена форма, сложни форми на прилагателните, диалектни думи или фонетични варианти. Друга типична черта е наличието на колебания в основната форма при имена от чужд произход, например в случаите, когато не е ясно какъв е граматичният род (ако има такъв) на топонима в изходния език (срв. *Южна Нахани* и *Южен Нахани*). Налице са и колебания при изписването с главни букви – особено при имена, производни от други, вече съществуващи имена, които на свой ред са образувани от нарицателни. Такъв е случаят с *Долен Плочник* (от вече съществуващото наименование *Плочник*). За да се определи дали и двете съставки се изписват с главна начална буква, или не, е необходимо да се уточни кой е източникът. Това невинаги е безпроблемно предвид разнородната писмена практика в интернет, от една страна, и липсата на пълна експертна кодификация – от друга. Не на последно място единиците от други езици, а и някои диалектни топоними може да имат непрозрачна лексикална и синтактична структура, например *Шиша Пангма*, *Тел Авив*, *Тал Афар* и дори наименования от други славянски езици като *Кутна Хора*, *Руда Силезка*. При подобни примери според случая е възможно разширяване на речника и най-вече приемане на обобщени граматични етикети за части от наименованията, които да служат за анотирането на непрозрачните компоненти на имената или изискват познания по конкретни езици или диалекти.

По аналогичен начин са описани всички близо 2600 географски наименования, като работата по обогатяването и прецизирането на описанието продължава. Флективните типове се основават на вече дефинирани типове на съставни лексикални единици не-имена чрез въвеждане на ограничения върху парадигмата, допълване и под. основно в посока на отразяване на морфофонемните редувания. Приложими са при описанието на единици от други области или от общата лексика, които се изменят по същия начин.

При описанието са използвани 95 подтипа природни, административни и други обекти (връх, планина, езеро, река; град, село, район, област), които са приписани еднозначно (при едно и също название, спадащо към различни типове, следва да се въведат отделни единици) и са проверени ръчно.

Речникът обхваща разнородни по произход и състав единици, включително единици с непрозрачна вътрешна структура (при които не може да се определи коя е главната част) и нефигуриращи в речника (не могат да се разпознаят от програмата за анотация).

Използвани са 134 флективни типа, като работата по обогатяването и прецизирането им е текуща задача.

Включените единици спадат към няколко структурни типа: съчетание от прилагателно и съществително (A N): *Троянски пролом*, *Видинско пръс-*

кало, *Тракийска низина* (65 флективни типа); съчетание от числително и съществително (B N): *Две тополи*, *Деветте извора* (7 флективни типа); съчетание от две съществителни, като към този тип спадат няколко структурни подтипа: *Ерма река*, *Баба Стана*, *Тодор Икономово*, *Чобан могила*, *Извор махала* (51 флективни типа); по-редки съчетания, представени от единични примери: съчетание от глагол в повелителна форма и съществително: *Плачи могила*; съчетание от предлог и съществително: *Под язовира.* Дефинирани са и типове, в които един или и двата компонента са неопределени (не е ясно към коя част на речта принадлежат и/или каква е синтактичната връзка между компонентите; не може да се определи родът и т.н.).

6. Заключение

Описанието и валидацията на единиците в речника могат да се използват за обогатяване на съществуващите правописни и тълковни речници и бази от данни с информация за парадигмата на съставните наименования. Както показват експериментите за други езици (Baziotis et al. 2023; Dimakis et al. 2024), речниковата информация има приложение и за усъвършенстването на предиктивните модели за езиково описание. Работата по граматичния речник продължава с попълване с подходящи единици и нови флективни типове.

Благодарности

Работата, чиито резултати са представени в настоящата статия, е извършена в рамките на проекта „Семантични ресурси и програми за обработка на езика (лексикално-семантични мрежи и езикови модели)“ (2023 – 2025) на Секцията по компютърна лингвистика на Института за български език „Проф. Любомир Андрейчин“ при Българската академия на науките.

БЕЛЕЖКИ / NOTES

[1] http://sgjp.pl/

[2] https://seelrc-iis.trinity.duke.edu/russdict/

[3] https://beron.mon.bg/

[4] Благодаря на анонимния рецензент, който обърна внимание на този факт и предостави примерите, включени по-горе.

ЛИТЕРАТУРА

Заимов 2012: *Заимов, Й.* Български водопис: Географско описание, строеж и произход на имената. Велико Търново, Университетско издателство „Св. св. Кирил и Методий“, т. 1.

Заимов 2012: *Заимов, Й.* Български водопис: Географско описание, строеж и произход на имената. Велико Търново, Университетско издателство „Св. св. Кирил и Методий“, т. 3.

Коева 1998: *Коева, Св.* Граматичен речник на българския език. Описание на концепцията за организацията на лингвистичните данни. – *Български език*, № 6, с. 49–58.

Тодорова, Лесева, Стоянова 2017: *Тодорова, М., Св. Лесева, Ив. Стоянова.* Речник на съставните думи в българския език – развитие и перспективи. – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“ (София, 15–16 май 2017 година)*. Т. 1, София, Институт за български език „Проф. Любомир Андрейчин“, с. 311–319.

Baziotis et al. 2023: *Baziotis, C., P. Mathur, E. Hasler*. Automatic evaluation and analysis of idioms in neural machine translation. – In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp 3682–3700, Dubrovnik, Croatia. Association for Computational Linguistics.

Courtois, Silberztein 1990: *Courtois, B., M. Silberztein.* Dictionnaires ´electroniques du francais. – *Langue francaise,* 87 (1), pp. 11–22.

Dimakis et al. 2024: *Dimakis, A. S. Markantonatou, A. Anastasopoulos*. Dictionary-aided translation for handling multi-word expressions in low-resource languages. – In: *Findings of the Association for Computational Linguistics (ACL) 2024*. Bangkok, Thailand and virtual meeting, Association for Computational Linguistics, pp. pp. 2588–2595.

Gross 1997: *Gross, M.* The construction of local grammars. – In: Roche, E., Y. Schabs (eds.). *Finite State Language Processing*, pp. 329–354.

Koeva, Stoyanova, Todorova, Leseva 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.

Sag et al. 2002: *Sag, I. A., T. Baldwin, F. Bond, A. A. Copestake, D. Flickinger.* Multiword expressions: A pain in the neck for NLP. – In: *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing (CICLing '02*). Springer-Verlag, pp. 1–15.

Savary et al. 2019: *Savary, A., S. Cordeiro, C. Ramisch*. Without lexicons, multiword expression identification will never fly: A position statement. – In: *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019*). Florence, Italy. Association for Computational Linguistics, pp. 79–91.

REFERENCES

Baziotis et al. 2023: *Baziotis, C., P. Mathur, E. Hasler*. Automatic evaluation and analysis of idioms in neural machine translation. – In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp 3682–3700, Dubrovnik, Croatia. Association for Computational Linguistics.

Courtois, Silberztein 1990: *Courtois, B., M. Silberztein.* Dictionnaires ´electroniques du francais. – *Langue francaise,* 87 (1), pp. 11–22.

Dimakis et al. 2024: *Dimakis, A. S. Markantonatou, A. Anastasopoulos*. Dictionary-aided translation for handling multi-word expressions in low-resource languages. – In: *Findings of the Association for Computational Linguistics (ACL) 2024*.

Bangkok, Thailand and virtual meeting, Association for Computational Linguistics, pp. pp. 2588–2595.

Gross 1997: *Gross, M.* The construction of local grammars. – In: Roche, E., Y. Schabs (eds.). *Finite State Language Processing*, pp. 329–354.

Koeva 1998: *Koeva, S.* Gramatichen rechnik na balgarskiya ezik. Opisanie na kontseptsiyata za organizatsiyata na lingvistichnite danni. – *Balgarski ezik*, N 6, s. 49–58.

Koeva, Stoyanova, Todorova, Leseva 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.

Sag et al. 2002: *Sag, I. A., T. Baldwin, F. Bond, A. A. Copestake, D. Flickinger.* Multiword expressions: A pain in the neck for NLP. – In: *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, CICLing '02*. Springer-Verlag, pp. 1–15.

Savary et al. 2019: *Savary, A., S. Cordeiro, C. Ramisch*. Without lexicons, multiword expression identification will never fly: A position statement. – In: *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*. Florence, Italy. Association for Computational Linguistics, pp. 79–91.

Todorova, Leseva, Stoyanova 2017: *Todorova, M., Sv. Leseva, Iv. Stoyanova*. Rechnik na sastavnite dumi v balgarskiya ezik – razvitie i perspektivi. – In: *Dokladi ot Mezhdunarodnata yubileyna konferentsiya na Instituta za balgarski ezik "Prof. Lyubomir Andreychin" (Sofia, 15–16 may 2017 godina)*. T. 1, Sofia, Institut za balgarski ezik "Prof. Lyubomir Andreychin", s. 311–319.

Zaimov 2012: *Zaimov, Y.* Balgarski vodopis. Geografsko opisanie, stroezh i proizhod na imenata. Veliko Tarnovo, St Cyril and St Methodius University Press, t. 1.

Zaimov 2021: *Zaimov, Y.* Balgarski vodopis. Geografsko opisanie, stroezh i proizhod na imenata. Sofia, Prof. Marin Drinov Publishing House, t. 3.

✉ *Гл. ас. д-р Светлозара Лесева*
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

✉ *Assist. Prof. Dr. Svetlozara Leseva*
Department of Computational Linguistics
Institute for Bulgarian Language "Prof. Lyubomir Andreychin"
Bulgarian Academy of Sciences
52 Shipchenski prohod, Bl. 17, Sofia 1113, Bulgaria