

## **ЧАСТИТЕ НА РЕЧТА В ИНИЦИАТИВАТА ЗА УНИВЕРСАЛНИТЕ ЗАВИСИМОСТИ: СТАТУС И ПОЛЕМИКИ**

**ПЕТЯ ОСЕНОВА**

СОФИЙСКИ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

*osenova@uni-sofia.bg*

Статията разглежда моделирането на частите на речта и техните характеристики – лексикални и граматически – в рамките на многоезиковата инициатива за Универсалните зависимости с оглед на българския език. Специално внимание се отделя на моделирането на спомагателните глаголи и детерминаторите. Заключават се, че от една страна, универсалният модел за представяне на морфологична информация е полезен за типологически съпоставяния между езиците, но от друга страна, понякога цената на съпоставимостта се заплаща с въвеждане на нетипични модели за представяне на езиковоспецифичните явления.

*Ключови думи:* части на речта; лексикални и граматически характеристики; български език; Универсални зависимости; езиковоспецифични явления; езиково моделиране

## **PARTS OF SPEECH WITHIN THE UNIVERSAL DEPENDENCIES INITIATIVE: STATUS AND DISCUSSION**

**PETYA OSENOVA**

SOFIA UNIVERSITY ST. KLIMENT OHRIDSKI

*osenova@uni-sofia.bg*

The paper focuses on the modelling of the parts of speech and their lexical and grammatical characteristics in the framework of the multilingual Universal Dependencies initiative with a view to the description of Bulgarian. Special focus is put on the modelling of auxiliaries and determiners. I conclude that, on the one hand, the Universal model is useful for typological comparisons among languages, while on the other, the price of rendering language structures comparable cross-lingually is sometimes paid by the adoption of non-typical models for representing language-specific phenomena.

*Keywords:* parts of speech; lexical and grammatical characteristics; Bulgarian; Universal Dependencies

Текущата класификация на частите на речта в инициативата за Универсалните зависимости\* претендира за универсалност по отношение на голяма група езици – както типологически сродни, така и типологически разнородни. Класификацията включва два типа информация – части на речта и

техните характеристики. Чрез комбинацията от двете могат да се представят различни езиковоспецифични особености. Но ако този начин се окаже недостатъчен, тогава има възможност да се добавят нови стойности в модела. Подобен морфологичен езиков модел не е нещо ново в лингвистиката. Новото обаче е, че той успешно се прилага върху голям брой езици. Морфологичен модел за българския е представен и в (Паскалева/Paskaleva 2007: 47). Авторката го разделя на три равнища: първо равнище – част на речта, второ равнище – лексикални подкатегории като род при съществителните имена и вид при глаголите, и трето равнище – граматични подкатегории като число при съществителните, прилагателните, глаголите или род при прилагателните.

Като всеки друг морфосинтактичен модел за естествените езици, и тази класификация поражда различни дискуссионни моменти. Те се отнасят до нейния вид и обхват; до избора на критерий или набор от критерии при определяне на частите на речта (морфологичен, синтактичен, семантичен или смесен подход); до отношенията между частите на речта и допълнителните техни лексикални и/или граматически характеристики.

Видът на класификацията отчита отворени и затворени класове думи, както и други класове.

По отношение на обхвата на класификацията се предлагат 17 части на речта, разпределени както следва: отворени части на речта (шест на брой), затворени части на речта (осем на брой) и други части на речта (три на брой). Включени са традиционните 10 части на речта (съществително име, прилагателно име, числително име, местоимение, глагол, наречие, предлог, съюз, частица, междуметие), но някои от тях са представени със своите подвидове (подчинителен и съчинителен съюз; спомагателен глагол; собствено име), а други са типологически окрупнени (адпозитив, който обхваща предлози и следлози). Трети са по-скоро морфосинтактични и варират в различните езици (например детерминаторът).

В групата на отворените класове са включени следните: прилагателно, наречие, съществително, съществително собствено, междуметие и глагол. Тук по-особеното е, че собственото име е отделна част на речта. Има само една функционална част в тази група – междуметието.

В групата на затворените класове са включени: адпозитив (предлог или следлог), спомагателен глагол, съчинителен съюз, детерминатор, числително име, частица, местоимение, подчинителен съюз. Тази група съдържа основно функционални думи. В нея се наблюдават следните интересни явления: отделянето на спомагателния глагол като отделна част на речта; отделянето на съюза на съчинителен и подчинителен; обединяването на предлога и следлога в обща част на речта; включването на детерминатора.

Групата *Други* съдържа: пунктуацията, специалните символи и чуждите думи (както и неразбираемите поредици от букви). Подобна група не е типична за традиционните лингвистични класификации, но е полезна, защото,

от една страна, пунктуацията също е значеща, а от друга – трябва да се предложи анализ на символите и чуждите думи, които се срещат в рамките на текста.

Трябва да се отбележи, че и в българската граматическа литература има представени класификации на частите на речта, които включват нетрадиционни членове. Такива са например в (Георгиев/Georgiev 1999: 13–14) предикативът (безлични неизменяеми думи, срещащи се предимно в предикативна употреба и като част от съставно сказуемо като *топло е, мъчно ми е*) и детерминативът (неизменяеми думи с логическо, модално или емоционално значение като *наистина, навярно, непременно*). Видно е, че добавянето на тези части на речта е продиктувано от синтактичната и/или прагматичната им употреба.

Друг български граматик, който задълбочено се занимава с рекласификацията на частите на речта по посока на по-адекватно представяне на тяхната функционална страна, е Константин Куцаров (вж. например Куцаров/Kutsarov 2019 и Куцаров/Kutsarov 2022). Интересно е да се отбележи също така, че в (Куцаров/Kutsarov 2011: 115–121) е представен обзор на вариациите на броя части на речта не само в българските граматика – включително в диахронен план, но и в граматиките на други славянски езици. Авторът се придържа към традиционните десет части на речта, които обаче йерархизира, разделяйки ги на основни (съществително, прилагателно, глагол, наречие) и неосновни, които, от своя страна, се разделят на следните видове: надстроечни (числително име и местоимение), несамостоятелни (предлог, съюз, частица) и специфични (междуметие).

Интересни в типологичен план в модела на Универсалните зависимости се оказват спомагателните глаголи (в отношението им към пълнозначните глаголи) и детерминаторите (в отношението им към местоименията). Спомагателният глагол в българския език е *съм* (с варианти *бъда* и *бивам*). Освен него обаче като спомагателни се разглеждат и частиците *да* и *ще*. Спомагателният глагол като част на речта не само спомага за образуването на глаголни форми (*доишъл си*), но има отношение и към копулата като синтактична връзка за образуване на предикативни конструкции с неглаголни компоненти (*учител съм*). Къде обаче попадат модалните глаголи от типа на *трябва, следва, може, мога*? В нашия модел в рамките на Универсалните зависимости те се разглеждат като пълнозначни (за разлика от английския, където са разгледани като спомагателни глаголи), защото добавят модална семантика към спомагателната си функция. Всъщност не е толкова лесно да се прецени къде попадат подобни групи глаголи поради факта, че притежават черти и от двата възможни класа – спомагателния и пълнозначния.

Невинаги е лесно да се направи разлика между местоимение и детерминатор особено ако те „съжителстват“ на едно и също равнище, в случая – морфологичното. В предложената класификация се следва подходът на Лайънз (Lyons 1991: 169), който предлага прецизиране на термина *pronoun* по

следния начин: *pro-noun* (место-съществително), *pro-adjective* (место-прилагателно) и т.н. По този начин в обхвата на местоимението се разбира само съществителното име, в обхвата на детерминатора се разбират не само артиклите, но и атрибутивните местоимения, а в обхвата на наречieto – адвербиалните местоимения. В българската традиция в рамките на местоимението влизат както заместващите съществителни, така и заместващите прилагателни и наречия (срв. в (Ницолова/Nitsolova 2008)).

Детерминаторът се дефинира като модификатор на съществителните, който определя тяхната референция в контекста. В езиците, в които артиклите са отделни думи, те, разбира се, се третираат като детерминатори. В българския език обаче определителният член е с афиксална природа и на морфологично равнище е част от думата. Така за детерминатори може да се разглеждат само местоименията в атрибутивна функция (*този, мой, какъв, някакъв, всякакъв* и т.н.). Следователно е направено и в текущия модел. Подобен подход обаче изкуствено третира един категориален признак при местоименията като отделна част на речта. Проблемен е фактът, че някои местоимения са омонимни в употребите си като съществително и прилагателно. Затова се налага подобна многозначност да се решава на синтактично равнище. Срв. *Някой те търси* и *Някой приятел може да помогне*.

Допълнителните лексикални и/или граматически характеристики съответно се влияят от степента на детайлност в частите на речта. Те са разделени на два типа: лексикални и флективни. Лексикалните включват: тип на местоимението, тип на числителното, посесивност, рефлексивност, чужда дума, абревиатура, правописна грешка. Типът местоимение дава възможност да се избере между артикъл (в българския би било приложимо към лексемата *един*, ако се следва идеята, поддържана от редица учени в нашето езикознание, че *един* изпълнява функцията и на неопределителен член), показателно местоимение (*този, тази, това*), емфатичен детерминатор (*Той сам го направи*), детерминатор за изненада (*Какъв ден!*), неопределително местоимение (*някой, някоя, някое*), въпросително местоимение (*кой, коя, кое*), отрицателно местоимение (*никой, никоя, никое*), лично (*аз, ти той*) или притежателно местоимение (*мой, твой, негов*), реципрочно местоимение (*един друг*), относително местоимение (*който, която, което*), обобщително местоимение (*всеки, всяка, всяко*). Както се вижда, в разглеждания модел границата между прономиналност и детерминативност е доста размыта, което не е изненадващо предвид факта, че детерминаторите като цяло имат местоименна природа, а някои местоимения се държат като детерминатори на синтактично равнище. В българския модел обаче не отделям специално емфатичния детерминатор, детерминатора за изненада и реципрочното местоимение, защото емфатичният детерминатор се разглежда като прилагателно име; детерминаторът за изненада е транспозиция на основната употреба на местоименията, а реципрочното местоимение се разглежда първо чрез отделните си части, а после като фраза с реципрочно значение.

Типът числително представя следните стойности: бройно числително (*едно, две, три*), редно числително (*първи, втори, трети*), мултипликативно числително (*двоен, троен, веднъж, дваж*), дробно числително (*две трети*), колективно числително (*два чифта*), дистрибутивно числително (*по един лев на дете*), числителни с обхват (*две-три, десет-двайсет*). В българския модел мултипликативните числителни са разгледани съответно като прилагателни (*двоен, троен*) или наречия (*веднъж*), т.е. традиционно е даден превес на синтактичния, а не на семантичния критерий. Дробното числително също е разгледано чрез своите съставни части – бройно и редно числително. И колективното числително в български се разглежда на равнището на фразата. Числителните с обхват, които в български са за приблизителност, се анализират като бройни числителни, а не самостоятелно. В модела няма дистрибутивни числителни.

Посесивността има стойност по подразбиране НЕ и експлицитна стойност ДА. Същото се отнася и за рефлексивността, за чуждата дума, абривиатурите и правописните грешки. Посесивността се отнася не само до притежателните местоимения, но и до притежателните прилагателни като в съчетанието *майчина сълза*. Рефлексивността се отнася до личните (*себе си, се, си*) и притежателните местоимения (*свой, своя, свое*). Останалите три характеристики маркират съответно наличието на елементи от чужд език в текста, абривиатури и правописни грешки. В българския модел обаче абривиатурите се анализират като съществителни собствени.

Флективните характеристики се разделят на два подтипа: номинални и вербални. Номиналните включват: род, одушевеност, тип на съществителното, число, падеж, определеност и степен. Родът е представен с трите си грамеми: мъжки, женски и среден. По-интересна е характеристиката одушевеност. Аз съм от езиковедите, които смятат, че определяща за българския език е опозицията *лице – нелице*, а не *одушевлено – неодушевлено*. Но в самия модел като маркери за одушевеност се смятат: акузативните форми при мъжки род на някои местоимения (*кого, някого, когото* и др.), количествените форми за лица *колцина, неколцина*, както и личните и притежателните местоимения в 1. и 2. лице – *аз, ние, ти, вие, мой, твой, наш, ваш*. След това се открояват одушевени същности, които не са лица, като например *два лъва*, и неодушевени, като например *две маси*. Типът на съществителното е характерен за определена група езици, какъвто например е банту, но не и за българския език. Категорията число има стойности за единствено (*молив*) и множествено (*моливи*) число при броимите същности. Освен това е добавена стойност за бройната форма като наследник на двойственото число с редица граматически и семантични ограничения (*два стола, три гълъба*), която е езиковоспецифична. Отделно се маркират и съществителните, които са *pluralia tantum*, но не са маркирани тези, които са *singularia tantum*. Категорията падеж отчита звателната форма при съществителните, когато тя е маркирана (*Госпожо! Господине!*). При местоименията са добавени стой-

ности за именителен, винителен и дателен падеж при личните местоимения и за винителен и дателен падеж при въпросителните и някои други видове. Дефинитността отчита като неопределени употребите с *един, една, едно* и нулевия маркер (*една жена, 0 жена*), а като определени – формите с определен член (*жената*). Категорията степен е маркирана с трите стойности – положителна (*добър*), сравнителна (*по-добър*) и суперлативна (*най-добър*).

Вербалните категории включват: тип на глагола, наклонение, време, вид, залог, свидетелственост, полярност, лице, учтивост и изключителност. Типът на глагола отразява за българския следните стойности: лична форма (*чета*), причастие (*чел, четен*), деепричастие (*четейки*). Не са отделени със специален таг отглаголните съществителни (*четене*). Наклонението за български включва три грамеми: изявително (*всички глаголни времена*), повелително (*подай, вземи, ела*) и условно (*бих дошъл*), което обаче е аналитично и затова на това равнище се разглежда чрез своите елементи. От деветте глаголни времена в нашия език само три са синтетични – сегашно, минало свършено и минало несвършено. Затова и само те са кодирани директно. Категорията вид има две стойности: свършен (*разбера*) и несвършен (*разбирам*). Няма експлицитна категория за двувидовите глаголи (*иронизирам*), макар че има място и за подобна възможност в морфологичния модел. Залогът има само две стойности: деятелен (*видях*) и страдателен (*видян съм*), макар че в българската лингвистична литература битуват още тризалоговите и четиризалоговите теории. В реалността обаче залозите се маркират върху глаголната форма, ако тя е синтетична. При аналитичните форми информацията се носи от отделните елементи. Най-трудно е със *се*-спрежението на страдателния залог поради факта, че елементът *се* е многозначен. Срв. *Книгата в момента се изучава от учениците* (страдателен залог), *Много се смях на тази случка* (деятелен залог), *Облякох се бързо* (възвратен залог – при три- и четиризалоговите теории). Свидетелствеността в момента не се маркира на това равнище в модела за българския език. За обусловеността на подобна стъпка от контекста и знанията ни за света, както и за корпусен поглед към тенденциите за съчетаване на свидетелски и несвидетелски стратегии в българския език вж. в (Tagromanova and Aleksova 2022). Полярността е характеристика, която противопоставя положителни думи с техните негирани съответствия. Напр. *добър – недобър*. Тази информация обаче също не се кодира в модела за българския език. Категорията лице има три стойности – за първо (*аз; летя*), второ (*ти; гледаш*) и трето (*той, тя то; върви*) лице. Трябва да се отбележи, че категорията е характерна за глаголите и местоименията.

Учтивостта и изключителността също не се маркират специално в българския модел. За учтивостта може да се направи маркиране според официалното и неофициалното общуване с използването на второличните форми *ти/Вие*. При категорията изключителност в българския няма разлика при

включване или изключване на слушателя спрямо множеството на говорителя. Изречението *Ние идваме* може да изразява както множеството *аз/ние* + *ти/вие*, така и множеството *аз/ние* + *той/тя/то/те*.

Въз основа на разсъжденията ми по-горе бих искала да изразя своята лична позиция по отношение на модела за представяне на частите на речта за българския език в контекста на Универсалните зависимости. Тъй като в идеите на Универсалните зависимости се запазват традиционните части на речта с малки размествания по отношение на специфичност или окрупняване, както и с малки добавки, сред които и връзки със синтаксиса, проектирането на информацията върху десетте части на речта за българския език е сравнително директно и безпроблемно. Тук бяха обсъдени някои от предизвикателствата.

В рамките на Универсалните зависимости едномерността на класификацията на частите на речта се компенсира и с комбинирането им със съответните лексикални и/или граматически характеристики. По отношение на нашата традиционна класификация подобна комбинация с лексикалните и/или граматическите характеристики също се описва подробно и задълбочено в граматиките на българския език, което дава възможност за построяване на различни формални езикови модели.

Със сигурност не съществува идеална класификация на частите на речта за даден език, камо ли за голям набор от езици. Но в стремежа ни към подобна класификация аз мисля, че десетте традиционни части на речта са много добра основа, над която може да се надгражда според целите ни и според перспективата, от която искаме да изучаваме езика.

В заключение може да се каже, че морфологични модели, които са изградени върху данни от много езици, са с доста богата класификационна схема, отчитаща взаимодействието между различните езикови равнища. Този факт, от една страна, подпомага моделирането на явленията в отделните езици, като го прави по-последователно и типологически съпоставимо, но от друга, неминуемо води до трудности при вземането на определени лингвистични решения и създава предпоставки за неточности или нетипични интерпретации в езиковоспецифичните модели.

#### БЕЛЕЖКА / NOTE

\* <https://universaldependencies.org/>

#### ЛИТЕРАТУРА

- Георгиев 1999: *Георгиев, Ст.* Морфология на българския книжовен език. Велико Търново, четвърто издание.
- Куцаров 2011: *Куцаров, Ив.* Към въпроса за частите на речта и тяхната класификация. – *Актуална морфология*. Пловдив, Университетско издателство „Паисий Хилендарски“.

- Куцаров 2019: *Куцаров, К.* Концепцията за частите на речта при Лев Владимирович Щерба. – *Българистични езиковедски четения. Материали от Международната конференция по случай 130-годишнината на Софийския университет „Св. Климент Охридски“*, 19–20 ноември 2018 г.
- Куцаров 2022: *Куцаров, К.* Българските лексемни класове и учението за частите на речта. Пловдив, ИК Колибри.
- Ницолова 2008: *Ницолова, Р.* Българска граматика. Морфология. София, Университетско издателство „Св. Климент Охридски“.
- Паскалева 2007: *Паскалева, Е.* Компютърна морфология. Ресурси и инструменти. София, ИПОИ – Българска академия на науките.
- Lyons 1991: *Lyons, J.* Natural Language and Universal grammar. Essays in Linguistic Theory. Cambridge – New York – Port Chester – Melbourne – Sydney, vol. 1.

#### REFERENCES

- Georgiev 1999: *Georgiev, St.* Morfologiya na balgarskiya knizhoven ezik. Veliko Tarnovo, chetvarto izdanie.
- Kutsarov 2011: *Kutsarov, Iv.* Kam vaprosa za chastite na rechta i tyahnata klasifikatsiya. – *Aktualna morfologiya*. Plovdiv, Universitetsko izdatelstvo “Paisiy Hilendarski”.
- Kutsarov 2019: *Kutsarov, K.* Kontseptsiyata za chastite na rechta pri Lev Vladimirovich Shcherba. – *Balgaristichni ezikovedski cheteniya. Materiali ot Mezhdunarodnata konferenciya po sluchay 130-godishninata na Sofiyskiya universitet “Sv. Kliment Ohridski”*, 19–20 noemvri 2018.
- Kutsarov 2019: *Kutsarov, K.* Balgarskite leksemni klasove i uchenieto za chastite na rechta. Plovdiv, IK Kolibri.
- Lyons 1991: *Lyons, J.* Natural Language and Universal grammar. Essays in Linguistic Theory. Cambridge – New York – Port Chester – Melbourne – Sydney, vol. 1.
- Nitsolova 2008: *Nitsolova, R.* Balgarska gramatika. Morfologiya. Universitetsko izdatelstvo “Sv. Kliment Ohridski”. Sofia.
- Paskaleva 2007: *Paskaleva, E.* Kompyutarna morfologiya. Resursi i instrumenti. Sofia, IPOI – Bulgarian Academy of Sciences.
- Tarpomanova and Aleksova 2022: *Tarpomanova, E., K. Aleksova.* Evidential strategies and grammatical marking in clauses governed by verba dicendi in Bulgarian. – *Proceedings of the Fifth International Conference Computational Linguistics in Bulgaria (CLIB 2022)*, Institute for Bulgarian language, Bulgarian Academy of Sciences, pp. 81–88.

✉ Проф. д-р Петя Осенова  
Факултет по славянски филологии  
Софийски университет „Св. Климент Охридски“  
бул. „Цар Освободител“ 15, 1504 София, България

✉ Prof. Petya Osenova, PhD  
Faculty of Slavic Studies, Sofia University St. Kliment Ohridski  
15 Tsar Osvoboditel, Sofia 1504, Bulgaria