

## РЕЦЕНЗИЯ

от проф. д-р Гургана Кирилова Дачева – СУ „Св. Кл. Охридски”,

за научните трудове и научните приноси

на доц. д-р Петя Начева Осенова,

кандидат в конкурс за професор по 2.1. Филология (морфология, синтаксис и корпусна лингвистика), ДВ, бр. 9 от 02.02. 2016 г.

### *1. Данни за кандидата:*

Петя Начева Осенова завърва висшето си образование в СУ „Св. Климент Охридски” – специалност чешка и българска филология, през 1995 г. През същата година завършва и втора специалност английски език и литература. От 1995 г. до 1996 г. работи като преподавател по английски език в училища „Фарос”. В периода 1996-1999 г. е редовен докторант по съвременен български език (морфология) в Института по български език към БАН. През 1999 защитава дисертация на тема „Функциониране на неопределителните местоимения в българския език”. В периода 1999/2000 г. е хоноруван асистент в Катедрата по български език, като през същата година печели асистентски конкурс по морфология и синтаксис и остава в Катедрата като щатен преподавател. През 2003 е старши асистент, през 2004 – главен, а през 2009 е избрана за доцент. От същата година до днес Петя Осенова води лекции по морфология и синтаксис; съвременен български език и формални граматика. Тя работи активно и в областта на формалната лингвистика и компютърната корпусна лингвистика.

Прави впечатление, че през годините (периода 2000-2016) Петя Осенова непрекъснато усъвършенства професионалната си квалификация и уменията си за преподаване, като участва в редица специализации, обучения и научноизследователски проекти. Най-впечатляваща и приносна е работата ѝ по моделирането на корпуса Бултрибанк, както и участието ѝ в различни международни проекти, свързани с прилагането на езиковите технологии в Холандия и Германия. През 2010 г. Петя Осенова печели стипендия Фулбрайт и прекарва 5 месеца в Станфордския университет при проф. Айван Саг, където разработва първата версия на българската ресурсна граматика в рамките на Опорната фразова граматика.

Доц. Осенова е съорганизатор на множество научни събития – Семинар по частична обработка на големи корпуси (2003, Ланкастър, Англия); Семинар по използване на синтактично анотирани корпуси (2005, Бирмингам, Англия); Пети семинар по свързани данни в лингвистиката (2016, Словения) и много други.

Петя Осенова редовно участва като рецензент на статии в престижни научни форуми по езикови технологии и компютърна лингвистика. Тя е изключително уважаван учен в международните професионални среди.

Доц. Осенова успешно предава своите знания и умения на по-младите изследователи, като досега има една успешно и в срок защитила докторантка, а в момента е научен ръководител на Александра Димова.

Понастоящем Петя Осенова е зам.-декан по научната и проектната дейност във Факултета по славянски филологии.

## **2. Научни трудове:**

**А. Публикации** общо за научноизследователската ѝ дейност- общо 4 **монографии**; **студии, статии и доклади** (от участие в регионални, национални и международни конференции), основно на български и английски език, от които 104 в областта на морфологията, синтаксиса и корпусната лингвистика; 3 **рецензии** за книги; 3 **съставителства**.

**Б. Публикации, представени за участие в конкурса за професор** - 1 **монография**, 43 **студии и статии в областта на морфологията, синтаксиса и корпусната лингвистика**.

*Монографията „Грамматическо моделиране на българския език (с оглед на обработката на естествен език). С., 2016”* е с приносен характер, като основният принос на труда е акцентът върху параметрите на граматическото моделиране на българския език с оглед на конкретни теоретични рамки, които не са били широко използвани досега. Показани са както специфичните, така и универсалните граници на явленията. Основно предимство е, че всеки модел е свързан с реален ресурс, който е достъпен за ползване.

Двата начина за представяне на морфосинтактичните и синтактичните явления (чрез имплементиране на ресурсна граматика и чрез аотиране на синтактичен корпус спрямо определена формална лингвистична теория, който след това се премоделира в други анотационни схеми) въвеждат съответно и два подхода към езиковия анализ (и синтез): единият е ‘отдолу-нагоре’, което означава, че се набляга върху комбинирането на данните, а другият е ‘отгоре-надолу’, което за авторката означава, че се измерват обхватът и точността на цялостно изградената граматическа рамка.

Описана е подробно Българската ресурсна граматика, която е разработена в рамките на съвременна лингвистична теория □Опорната фразова граматика. В своя първи вариант тя покрива основното количество морфологични правила и синтактични явления. За целите на сравнението граматиката е синхронизирана с общ граматичен модел (матрица) и е оценена върху преведеното множество от изречения на този общ модел. Моделирането в Ресурната граматика показва най-пълно и реалистично взаимодействието между езиковите равнища (морфология, синтаксис, семантика) и

между ресурсните компоненти (речник, йерархия на явленията и граматически правила).

Друг основен принос в труда е дискутирането на конституентни и депendentни анализи за българския език, което обогатява перспективите на езиковото моделиране. Всъщност Граматика на зависимостите, или още Депендентната теория, не е прилагана широко за българския език. Представени са два депendentни модела за българския синтактичен корпус Бултрибанк. Единият е наречен стандартизиран, защото преносът на релации следва дословно оригиналната анотационна схема. Този пренос показва обаче каква част от лингвистичната информация е експлицирана на морфологично и на синтактично равнище. Другият модел е наречен универсален, тъй като е свързан с общ модел, който трябва да бъде следван, защото потенциално има за цел да обхване явленията във всички езици. Тъй като, подобно на Ресурсната граматика, има предварително зададена схема, отчетливо изпъкват разликите в моделите и в езиковоспецифичните характеристики. И в Ресурсната граматика, и в универсалния депendentен модел се приема семантичен подход към описанието на явленията за разлика от оригиналния синтактичен ресурс Бултрибанк, при който преобладава морфосинтактичният. Важен принос в работата е моделирането на българския език в семантично ориентирани рамки, което е предизвикателство за флективната му граматическа природа.

Друг приносен момент е разглеждането на понятието *опора* както в различните имплементирани модели, така и в теоретичен план. Представени са тестовете за определяне на опората, видовете реализации на опората, равнището на нейното представяне, областта ѝ на действие, като са дискутирани особеностите ѝ в българския език.

Представените модели на лингвистично описание, както и самите езикови ресурси (граматика и корпуси), допринасят не само за обективното типологическо сравнение на българския език с други езици, но и за подобряване на автоматичната обработка на българския език в морфосинтактичен, синтактичен и семантичен план.

Монографията представя Петя Осенова в изключително значима професионална светлина, показва нейните способности да анализира и систематизира както практическа, така и теоретична информация. Приносът на авторката за корпусната лингвистика в България, както и в световен мащаб е значителен и неоспорим. Тя е учен с висок престиж и авторитет в научните среди.

*Статиите, свързани с морфологията, са 9 (2, 6, 9, 11, 19, 25, 26, 27, 29 от списъка с публикациите) за първи път въвеждат представянето на т.нар. многокомпонентни думи (multiwords) в българския език чрез описание с катена. Приносът на изследванията е в представяне на взаимодействието между поведението на тези думи в речника и текста чрез катена. В речника се указват местата, където катената може да се разкъса или разшири. От морфологична гледна точка се дават основните форми на елементите на многокомпонентните думи, възможните граматически характеристики (ако има изменяеми елементи) и информация при кои части на речта са възможни разширения*

или промяна на граматическите характеристики. Приносите на статиите са в следните аспекти: сравнението между сложни глаголи и сложни съществителни по отношение на реализация на вътрешни и външни аргументи; подготовката на данните (корпус и правила) за тестване на морфологичния анотатор и също така при анализа на грешките след действието на автоматичните модули; описанието на типовете многозначности, честотата на срещане на най-честите думи в корпус, както и факторите, които влияят на избора на правилен морфосинтактичен етикет; анализаране на грешките при морфосинтактичния анализ с оглед на лематизацията (привеждане на словоформата в основната форма); описание на основните типове морфологични многозначности от перспективата на отношението между лексема и словоформа на основата на морфологично анотиран корпус.

*Статиите, свързани със синтаксиса, са 16 (2, 5, 6, 7, 8, 9, 11, 20, 23, 25, 28, 32, 33, 34, 36, 39 от списъка с публикациите).* Част от статиите описват въпроси, които се отнасят до синтактичния ресурс Бултрибанк; друга част са свързани отново основно с катена и функционирането ѝ; трета част се отнасят към монографията и показват различните етапи, през които в годините са се движели научните интереси на доц. Осенова. Процесът на изследване в различните аспекти на проблематиката показват научното израстване на Петя Осенова в теоретичен и приложен аспект. В част от статиите се вижда процеса на изграждане на Новата рамка на Бултрибанк, която се поставя в среда на сравнимост с други езици и дава възможност за сравняване на автоматични морфологични и синтактични анализатори, обучени върху ресурса, с автоматичните средства за други езици. Описани са подходите при конвертирането. Те включват автоматичен модул, който използва правила, и ръчен модул, при който се налага санкция от специалист в случаите на по-трудните езикови явления (координация, апозиция и др.) или разминаване на анотационните схеми (напр. при Бултрибанк схемата е по-морфологично и формално насочена, докато при универсалните депендентности е по-синтактично и семантично изградена). Приносите в тази област са свързани със следните аспекти: анотирането на синтактичния корпус Бултрибанк със семантични роли върху аргументите на предиката – подлог, пряк и непряк обект; предварителна работа по ресурса, която се изразява в анотиране на текстовете със значения от лексикалната база и извличане на валентен речник; използване на лексикографските класове на глаголите (когнитивни глаголи, глаголи за движение, прецепция, притежание и др.); създаване на стратегия на анотиране; извличане на валентен речник от синтактично анализирания корпус Бултрибанк; представяне на анотационна схема за българско-английски паралелен синтактичен корпус с използване на хибриден метод за българския език; въвеждането на нова теоретична рамка за граматическо моделиране на българския език, която не е била прилагана до момента; описване на параметрите на локализиране на идеята за ресурсна граматика спрямо българския език.

*Статиите, свързани с корпусната лингвистика, са 20 (3, 4, 10, 12, 14, 16, 17, 18, 19, 21, 22, 24, 31, 35, 37, 38, 40, 42, 43, 44).* Нито една от статиите в този раздел не е директно свързана с тематиката в монографията. Статиите покриват следните области:

семантично аотиране; различни видове автоматична обработка на български текстове; корпуси и корпусно базирани изследвания и машинен превод. Приносът на доц. Осенова е в извличане на синтактични релации от Бултрибанк и интерпретиране на резултатите; създаването на ресурси за българския език от гледна точка на комбиниране на синтаксис, семантика и знания за света. Аотирани са пълнозначните части на речта: съществителни, глаголи, наречия и прилагателни. Една от статиите описва експерименти, свързани с модул, който да комбинира решаването на три задачи едновременно – морфосинтактично тагиране, депендентен парсинг и кореферентно свързване. Принос е не само комбинирането на тези три задачи за решаване в едно, но и фактът, че самият модел е езиково независим. Комбинираният модел постига подобни резултати на тези, които се получават от най-добрия морфологичен тагер за българския език, и по-високи резултати от тези на синтактичния парсер, включен в интегрирания модул за обработка на български текстове. Приносът на доц. Осенова е в определянето на морфосинтактичните характеристики, които се използват в модела.

Особено интересна е статия 24, която е статия програма, както основателно я е нарекла авторката. Тя представя инфраструктура за езикови ресурси и технологии. По-точно казано, описан е основният, базисният пакет от ресурси и технологии за българския език. Първо е въведена основната идеология на международната инициатива за езикови ресурси и технологии CLARIN. След това са описани и самите ресурси за българския език (корпуси, речници, списъци с имена). Накрая са представени и езиковите технологии (сегментатор, средство за разпознаване на собствени имена, морфологичен анализатор, средство за снемане на морфологичната многозначност, лематизатор, чънкер (или още частични граматика) и парсери). Авторката прави извода, че българският език не е сред езиците, които са ‘ресурсно бедни’ и усилията трябва да се насочат към интегриране на данни и технологии, тяхното адаптиране за различни групи потребители и задачи.

В друга група статии се разискват проблеми, свързани с автоматичния машинен превод; представя се модел за машинен превод от български към английски и от английски към български. Лингвистичното знание, което включва морфосинтактична и синтактична информация, както и частично семантична информация, е кодирано като фактори в системата за машинен превод Моузес.

### ***3. Цитирания***

От справката с цитиранията на Петя Осенова се вижда, че общо **библиографските цитирания** са 183, общо **индексираните публикации** са 34. Това показва, че авторката се ползва с професионален авторитет в научните среди, а изследванията ѝ са познати и се използват от колегите ѝ. Трябва да отчетем и факта, че по мнението на студентите Петя Осенова поднася интересни и практически насочени лекции, което няма как да открием сред библиографските цитирания, но е не по-малко важно.

### ***4. Преподавателска работа***

Доц. Петя Осенова преподава в бакалавърски и магистърски програми в СУ „Св. Кл. Охридски” по: морфология и синтаксис (бакалаври); съвременен български език и формални граматика на бакалаври и магистри.

### ***Заключение***

Идеите и научните приноси, реализирани в публикациите на Петя Осенова, единствен кандидат в конкурса за професор по 2.1. Филология (морфология, синтаксис и корпусна лингвистика), ДВ, бр. 9 от 02.02.2016 г., сериозните ѝ научни и професионални качества, ми дават основание убедено да подкрепя избора на Петя Осенова в конкурса за професор и смятам, че ще защити достойно както своя професионализъм, така и научния авторитет на Катедрата по български език, на Факултета по славянски филологии и на СУ „Св.Кл.Охридски”.

30.05.2016 г.

проф. д-р Гергана Дачева