

Рецензия

от доц. д-р Мария Стамболиева

Департамент „Англицистика“ на Нов български университет

по конкурс за заемане на академичната длъжност „професор“ в професионалното направление 2.1 Филология (български език – морфология, синтаксис и корпусна лингвистика), обявен от СУ „Св. Климент Охридски“, ДВ бр. 9 от 2.02.2016 г.

Доцент д-р Петя Начева Осенова е единствен кандидат в конкурса. Тя е редовен преподавател в Катедра „Български език“ на СУ „Св. Климент Охридски“ от академичната 1999/2000 г., а от 2009 г. е доцент, лектор по морфология, синтаксис, формални граматика и съвременен български език. От представената автобиографична справка, от справките за цитиранията, приносите, публикациите след присъждането на академичната длъжност „доцент“ веднага става ясно, че доц. Осенова отговаря на всички формални изисквания, определени в Правилника на СУ „Св. Климент Охридски“, за заемане на академичната длъжност „професор“

Освен като лектор по морфология, синтаксис и съвременен български език, доц. Осенова от години работи и в областта на формалната лингвистика и компютърната корпусна лингвистика в сътрудничество с учени от Лабораторията за лингвистично моделиране към ИИКТ, БАН, на първо място – с доц. д-р Кирил Симов и неговия екип. Ако може да се съди по научните ѝ изяви и публикациите, именно на това сътрудничество доц. Осенова в най-голяма степен дължи професионалното си израстване като добре известен в професионалните среди изследовател и желан партньор в научно-приложни проекти и разработки. От представените за конкурса 43 статии, само 7 не са в съавторство с К. Симов и други членове на ЛЛМ; на описанието на проблематиката, свързана с работата ѝ по проектите на ЛЛМ Бултрибанк, CLEF, CLARIN и др., са посветени най-приносните ѝ разработки. Организационната и проектна дейност на доц. Осенова също, с много малко изключения, е обвързана със сътрудничеството ѝ с ИИКТ.

Почти без изключение, публикациите на П. Осенова са значими теми и проблеми, свързани с формалното представяне на данни от българския език, предимно самостоятелно, но и в съпоставка, и с оглед на автоматичната

обработка на езика. Трудно е да се отделят статиите, посветени на компютърната корпусна лингвистика от статиите, посветени на проблемите на формалното описание на морфологията или синтаксиса, но ще следвам реда, в който тя самата ги представя в авторската си справка.

На морфологични проблеми, свързани с обработката на електронни корпуси, са посветени 9 статии. Част от тях описват участието на П. Осенова в проекти за автоматичен морфологичен анализ. Като лингвист, нейният принос е в подготовката на материала, типовете морфологична многозначност, анализа на сложни глаголи и съществителни в българския език. По-особено внимание заслужава статията „Синтаксисът на думите“, в която са изнесени интересни данни за съотношението между вербалност и номинативност при аналитичното и синтетично реализиране на аргументите в сложни думи като *ръкомахам* (от *махам с ръка*) или *снеговалеж* (от *Вали сняг*). Резултатите от тази статия са интегрирани в няколко разработки на доц. Осенова и доц. Симов върху морфосинтактичните единства.

Морфосинтактични са проблемите, които поставят пред автоматичния анализ т.н. комплексни лексеми (Multiple Word Lexemes), наречени от П. Осенова „многокомпонентни думи“. Четирите статии, посветени на тях, лично за мен бяха особено интересни, тъй като преди 20 години, в рамките на проекта GLOSSER, участвах в първите опити за класификация и формално описание на комплексни лексеми. Макар че за този период самата класификация, публикувана от Айвън Заг през 2002 г., не е претърпяла съществено развитие, формалното им представяне като подструктури, поддървета „дървенца“ (treelets) във вид на катени е сравнително нова, а П. Осенова и К. Симов първи прилагат анализа на О' Грейди (1998) върху български материал. Прилагането на този анализ е оправдан за двете групи комплексни лексеми, в които е възможна вариация във формата или реда на компонентите – полуфиксираните и синтактично гъбкавите. Напълно оправдано, в Бултрибанк фиксираните лексеми получават лексикален анализ. Интересна е идеята за използване на формалното представяне на катените в анализа на клитиките. Подобен подход Симов и Осенова използват и в анализа на сложни думи, като следват идеите на Грос (2010) и (2011) за катена на морфи. Въвеждането на еднотипно формално представяне на словосъчетания и сложни думи дава възможност за по-елегантно формулиране на трансформациите между тях. Ненужна е дългата бележка под линия за сложните лексеми в българския език в началото на статията “Modeling Lexicon-Syntax Interaction with Catenae” – такива сложни

лексеми има и в много други езици, в т.ч. английски. И едно терминологично уточнение: в авторската справка П. Осенова превежда „light verb structures“ като „леки глаголни конструкции“. Лекото тук не е конструкцията, а глаголят – това са конструкции с лек (десемантизиран) глагол, които са особено характерни за английския език.

В началото на съвместната си работа П. Осенова и К. Симов възприеха синтактичния модел на ОФГ, който за първи път беше приложен върху български материал в края на 80-те и началото на 90-те г. от сътрудниците на ЛЛМ – Таня Августинова и Карел Олива. П. Осенова и К. Симов обаче първи създадоха цялостно, комплексно представяне на голям корпус от текстове на български език в този формално-синтактичен модел. През последните години те се преориентират към депendentния синтаксис, пионер на който у нас също е сътрудник (и съосновател) на ЛЛМ – покойната проф. Елена Паскалева. Още през 80-те години Е. Паскалева правеше успешни опити за прехвърляне на лингвистична информация от конституентно към депendentно дърво. П. Осенова и К. Симов нито просто повтарят, нито просто продължават експеримента на Елена Паскалева. От една страна, те първи разработват в дълбочина преходите от конкретен конституентен към конкретни депendentни модели; от друга, те не провеждат просто експеримент, а прехвърлят богата лингвистична информация върху ново представяне на огромен езиков материал. На прекодирането на оригиналния корпус Бултрибанк в наскоро възприетия от Симов и Осенова универсален депendentен модел са посветени някои от най-интересните статии на Петя Осенова. Благодарение на новата рамка, Бултрибанк, която беше в среда на сравнимост с други синтактично аотирани корпуси в модела на ОФГ, вече е в среда на сравнимост с много повече корпуси и езици. В статията „Universalizing BulTreeBank: A Linguistic Tale about Glocalisation“ са описани подходите при конвертирането. Много интересен е и преходът от по-традиционни депendentни модели към универсалните депendentности, в които се забелязват по-интуитивни и мотивирани представяния, по-малко ад хок решения.

Статията “Semantic Role Annotation in BulTreeBank” представя работата на Петя Осенова по обогатяването на аргументите на предикатите в Бултрибанк със семантични роли. “A Treebank-Driven Creation of an Onto-Valent Verb Lexicon for Bulgarian” е процедура за извличане на валентен речник от Бултрибанк. В „Syntactic-Semantic Treebank for Domain Ontology Creation” се описва методология за създаване на синтактичен ресурс в определена

тематична област. Това е чудесна разработка с голяма приложна значимост, която с благодарност в момента ползваме в НБУ като насока за изработването на материали за електронно обучението по специализиран чужд език.

Хибридният метод, използван за представяне на анотационната схема за българско-английския паралелен синтактичен корпус (с изграждане на семантични структури на основата на информация от депendentни анализи) изглежда много обещаващ, но е трудно да се преценят възможностите му при толкова оскъдно представен материал (*кучето на Браун/Brown's dog*). Впрочем, за мое голямо съжаление, повечето представени статии са по-скоро описания на проекти, процедури и методологии, отколкото конкретни анализи на езиков материал.

Освен в създаването на корпуси като лингвиста в тандема с доц. Симов, П. Осенова участва и в множество приложения на тези корпуси: извличане на синтактични отношения, експерименти в областта на автоматичния превод и др.

Спирам се на монографията на доц. Осенова „Грамматическо моделиране на българския език с оглед на автоматичната обработка на естествен език“ в края на рецензията, тъй като тя обобщава работата на авторката по проекта Бултрибанк и ресурсните граматика и прехвърлянето на лингвистична информация от ОФГ към dependentни модели. Тук, разбира се, поради поголемия обем, проблемите са поставени в по-широк контекст, отколкото в свързаните с темата статии, и са разгледани с повече внимание към детайла. Напълно съм съгласна с думите на ПО, че „без отчитане на езиковите факти не може да има добър анализатор на естествен език“, но в текста намирам и някои анализи, които трудно примирявам с езиковите факти, а и не винаги разбирам, защото не са съпроводени с коментар. Ето някои от тях:

с. 63: „Спомагателните глаголи включват: глагола *съм*, частицата за бъдеще време *ще*, формата за бъдеще време *циях* и неличната форма *да*“. В известните ми формално-синтактични модели спомагателните глаголи в аналитични глаголни форми, както и частиците, които имат отношение към категориите време/лице/число, действително попадат под възел на спомагателни елементи AUX или INFL, но досега не съм попадала на анализ, в който те се определят като спомагателни глаголи.

с. 66: „Копулата в качеството си на спомагателен глагол образува синтактични съчетания (...)“ и още, на с. 67: „Копулата няма своя семантика (...) Копулата

дава информация само за времето, вида и т.н.“ На първо място, не виждам с какво може да спечели автоматичният анализ на естествен език от единен анализ на граматически маркер и глагол-връзка (който при това на свой ред може да се появи в аналитична форма за времева отнесеност, аспект или залог!) Освен това, тъй като докторската ми дисертация беше посветена на съпоставката на *to be* и *съм* с материал от голям преводен корпус, по никакъв начин не мога да се съглася с твърдението, че глаголт-връзка няма семантика. В системата на езика, „немаркиран“ не е „празен“.

с. 90: „В българската лингвистична литература формата *да* също се разглежда като комплементизатор. В настоящия модел обаче да се анализира само като нелична глаголна форма, която взема глагола като свой комплемент и наследява всички негови комплементи“. Това твърдение е неточно в няколко отношения. В българската лингвистична литература се прави разлика, от една страна, между *че/дали/ако* и пр. като комплементизатори и, от друга – *да*, като маркер за неутрализиране на категориите (време)/лице/число в INFL; а за това, глагол ли е *да* или частица, по-горе вече стана дума.

с. 92. „Отрицателната частица *не* се разглежда като спомагателен глагол, който взема комплементи.“ – тук коментарът ми се аналогичен на този за с. 63.

с. 170. В малко вероятния пример *Тя влезе бързо и усмихната* проблемът не е в това, че има съчинително свързване на фраза на прилагателното и фраза на наречието, а 1/ че в съчинителна връзка се намират адюнкт и комплемент и 2/ че комплементът следва адюнкта.

с. 171. Комплексът от аргументи за анализиране на детерминатора на именните фрази като тяхна опора не се свежда единствено до началното им положение във фразата.

Доц. Осенова положително има сериозни аргументи и убедителни отговори на поставените по-горе въпроси. Поставям ги тук, защото бих искала да чуя тези аргументи и отговори, а не ги намирам в текста на монографията.

Възражения буди у мен използваната от доц. Осенова лингвистична терминология, която нарушава една дълга традиция в българското езикознание. Не разбирам защо езиковед от катедра „Български език“ на голям български университет използва термини-чуждици като: *координация* (вм. *съчинително свързване/съчинителна връзка*), *субект* вм. *подлог*, *обект* вм. *допълнение* (да не говорим за *директен/пряк обект* и *недиректен обект* вм.

пряко и непряко допълнение), транзитивен/интранзитивен в.м. *преходен/непреходен, прономинални местоимения, акузативна/посесивна* клитика, *дативен предлог, вербална фраза, нумерална релация, инициална опора*. Искрено се надявам не това да е терминологията, която изучават студентите.

Направените по-горе бележки се отнасят до ограничен брой конкретни анализи и отделни термини и нямат отношение към общата ми оценка за приноса на доц. Петя Осенова към съвременната лингвистика в България и Европа – за мен този принос е безспорен. П. Осенова е най-яркият и успешен представител, сред езиковедите, на компютърната корпусна лингвистика у нас. В сътрудничество с доц. Кирил Симов, тя не просто продължи, а доведе до успешен край редица проекти и начинания на Лабораторията за лингвистично моделиране, с които тази лаборатория започна дейността си в средата на 80-те години. Повечето от нас, първите ѝ членове, особено от „филологическата“ група, създадохме предимно модели, протитипи, фрагменти. Благодарение на своето трудолюбие, последователност, упоритост, на солидната си лингвистична подготовка, както и на уменията си за работа в екип, Петя Осенова в сътрудничество с доц. Кирил Симов и екипа на ЛЛМ стана създател на първия представителен корпус на българския език, със статут на национален корпус, на първата цялостна система за автоматична обработка на българския език, на множество уникални за страната ни приложения. Не подценявам останалите ѝ разработки, напр. работата ѝ върху политически текстове, статиите ѝ и курсовете ѝ, посветени на работата със студентите, но за мен няма съмнение, че Петя Осенова вече е записала името си в историята на българското езикознание именно като един от най-успешните ни изследователи в областта на формалната и компютърна лингвистика. Без колебание ще гласувам за присъждането на доц. Петя Начева Осенова на академичната длъжност „професор“ в професионалното направление 2.1 Филология (български език – морфология, синтаксис и корпусна лингвистика).

София, 1.06.2016 г.

Член на научното жури:

доц. д-р Мария Стамболиева