

СТАНОВИЩЕ

по конкурс за заемане на академичната длъжност доцент в направление 2.1. Филология, научна специалност Общо и сравнително езикознание за нуждите на Секцията за компютърна лингвистика в Института за български език „Проф. Л. Андрейчин“, съгласно обявата в ДВ 102/23.12.2022 г.,
с кандидат: гл. ас. д-р Цветана Иванова Димитрова.

Изготвено от: доц. д-р Милена Петрова Добрева, Институт по математика и информатика – БАН

1. Обща характеристика на представените материали

Кандидатката е представила една монография и 22 статии и студии (14 в областта на езиковите ресурси и 8 в областта на историческите корпуси). От тези 22 публикации, 7 са статии в списания, 4 са студии в поредици и 11 са доклади в пълен текст в сборници на конференции с рецензиране и международна видимост.

От представените за конкурса публикации, осем са видими в SCOPUS. Това са 2/3 от всички публикации на кандидатката, които са видими в тази база от данни, което показва стабилно увеличаване на присъствието на кандидатката в публикации с международни високи наукометрични показатели. Забелязани са 125 цитирания на трудовете на кандидатката.

Документацията за участие в конкурса включва и справка за участията в национални и международни проекти (общо 14 на брой), справка за участието в научни форуми (34 на брой), документи от последната атестация на кандидатката за периода 2017-2020 г., както и грамота от една награда и потвърждение за визита като гост-изследовател в Университета Ка' Фоскари във Венеция, Италия. Предоставена е и справка за приносите на кандидатката и нейна биография, както и пълните текстове на всички публикации по конкурса според изискванията.

2. Обща характеристика на научната, научно-приложната и педагогическата дейност на кандидата. Участие в научни проекти и ръководство на проекти

Основните и дълготрайни интереси на кандидатката са в областта на компютърната лингвистика; яркото ѝ присъствие още от началните години на научния ѝ път в тази област е отбелязано с присъждането на награда за млад учен на името на

Марин Дринов пред 2011 г. само три години след защитата на дисертационния ѝ труд „Старобългарската именна фраза: към анотационна спецификация“ в Норвежкия университет за наука и технологии, Трондхайм.

Много удовлетворително е да се наблюдава развитието на кандидатката и добиването на зрялост в тази област със задълбочаване на работата в две направления, проблематиката около подготовката и анализа на исторически корпуси и по-широката област на езиковите ресурси и компютърното моделиране на езика.

В областта на подготовката на исторически корпуси кандидатката съчетава работа по формализирани описания на езикови явления с практическа приложимост в Българския национален корпус и в WordNet за български език.

Няма данни за подготвени дипломанти и докторанти, но има сериозен принос към научни проекти в сътрудничество с екипи от България и чужди страни.

В периода 2012-2023 г. кандидатката е участвала в 7 проекта (на един от тези проекти е ръководител) с финансиране от различни международни програми, както и в 7 проекта с финансиране от български национални и институционални изследователски програми.

В документите по този конкурс е особено видим приносът на кандидатката към работата на проектите „Семантична класификация на прилагателните в Българския УърдНет“ (финансиране от ДФНП) и „Семантична мрежа с широк спектър от семантични релации“ (ДН 10-3 / 14.12.2016), финансиран от Фонд „Научни изследвания“ (ФНИ) – част от представените публикации представят резултати, получени в рамките на тези проекти.

Кандидатката има поредица от участия в международни и национални научни форуми. През 2022 г. е била гост-изследовател във Венецианския университет Ка' Фоскари. От 2020 г. изнася и лекционен курс по Корпусна лингвистика в Магистърска програма „Компютърна лингвистика. Интернет технологии в хуманитаристиката“ на Факултета по славянски филологии към Софийския университет „Св. Климент Охридски“.

Профилът на кандидатката показва добро съчетание на проектно-изследователска, публикационна и преподавателска активност. Тя е припознат в международната колегия изследовател с участие в програмни комитети на форуми в областта на компютърната лингвистика.

3. Аналитична преценка на основните научни и/или научно-приложни приноси на кандидата.

В обширното поле на компютърната лингвистика Цв. Димитрова насочва своите усилия на няколко взаимно свързани и допълващи се теми с голяма приложимост при разработката на цифрови ресурси за българския език.

Част от работата ѝ е посветена **на изследване на клитиките в различни периоди от развитието на българския език**. Нейните изследвания се базират върху обстойни наблюдения над ексцерпирани примери от старобългарския, среднобългарския и ранния новобългарски периоди, съвременни литературни примери и появи в диалектна и разговорна реч. Събирането на материали от толкова източници само по себе си е предизвикателство и една особено добра практика е, че те са публично достъпни и могат да се използват и за целите на други изследвания в духа на отворената наука – примерите у нас за отворена наука като цяло в областта на хуманитаристиката не са широко застъпени. Работата на д-р Димитрова в областта на анализа на клитичните елементи е представена в монографията (1) и публикациите 16, 17, 18, 19, 20, които съвкупно дават един цялостен и по-задълбочен поглед върху употребата на клитиките.

Около 1/3 от публикациите на д-р Димитрова са посветени на проблеми около **организацията на данните и структурирането на различни езикови ресурси** с принос към добавянето на езиков материал, отразяващ съвременния български език в WordNet, публикации 2, 3, 4, 5; Българско-английският паралелен корпус със съотнесени (прости) изречения (BulEnAC) – публикация 13), както и структурирането на езиков материал, който отразява развитието на българския език в различни периоди от време (публикации 22 и 23). Съчетаването на теоретична и практическа насоченост в работата на кандидатката е едно от достоинства на нейната работа, която помага и за развиване на методологията при разработка на езикови ресурси и дава солидни практически резултати.

Кандидатката също изследва в своята работа проблемите около **представянето на езиков материал при създаването и развитието на езикови ресурси** (тя има особено голям принос към развитие на Българския WordNet и Българския национален корпус). В това направление тя участва активно в научноизследователски проекти.

В проекта „Семантична класификация на прилагателните в Българския УърдНет“ Цв. Димитрова допринася с научно ръководство, подготвяне на формализми за описание на данните и подробна аргументация на принципите на класифициране на няколко нива; тези постижения са представени в публикации 6 – 8.

В проекта „Семантична мрежа с широк спектър от семантични релации“ (ДН 10-3 / 14.12.2016), финансиран от Фонд “Научни изследвания” (ФНИ) кандидатката работи по формулирането и приписването на допълнителни семантични релации между съществителните (10) и съществителните и глаголите (9), както и включването на причастия в синонимни множества на прилагателни имена WordNet за български език (11). Формулирани са и стъпки за проверка на валидността на приписаните семантични типове по отношение на съчетаемостта със съществуващите семантични класове в WordNet за български език (12).

Работата на Цв. Димитрова показва едно сравнително рядко срещано у нас съчетание между отличното познаване на традиционните методики за анализ на развитието на езика със способността да се разработват формални езикови модели, описващи езикови явления в различни езикови ресурси с голям обхват на въвежданите данни. В добавка кандидатката показва отношение и принос в приложението на принципите на отворената наука, което е особено полезно за разширяването на обхвата и използването в изследвания и на други учени на наличните ресурси.

Прави впечатление, че кандидатката показва отлични резултати и при самостоятелни изследвания, и в работата с разнообразни екипи. Това е друг признак за научна зрялост и способност за собствени постижения, съчетана с успешното ръководене на и принос към екипна работа.

4. Критични бележки и препоръки по представените трудове

Лично за мен са особено интересни работите на кандидатката в областта на историческите корпуси от старобългарски текстове (16, 23). Прави впечатление доброто познаване на наличните корпуси и са много полезни съпоставките на техните подходи и елементи. Това, което би помогнало за един холистичен подход е и включването на наблюдения върху това за отговорите на какви научни въпроси се използват различните налични корпуси. Този аспект не е засегнат до момента и може да се обмисли как да се добави към работата в областта на историческите корпуси в съчетание и с изясняване на това как трябва да е анотиран един корпус, за да е максимално полезен при обучението на ученици и студенти на старобългарски език.

В работите на кандидатката са отбелязани различните подходи при подбора на това какви елементи да се анотират, но тези избори обикновено са функция от намеренията на авторите на корпуса този ресурс да се ползва за определен тип изследвания. Би било интересно да се свържат структурите на анотиране с намеренията на създателите и изследователските въпроси.

Намирам също така за много благодатен за разширяване на изследванията и въпроса дали може един исторически текст да се анотира с такъв обхват на анотациите,

че да служи за различни изследователски цели. Би могло да се проучи и дали методите на машинното обучение и изкуствения интелект не могат да послужат за допълващо тагиране на съществуващи ресурси.

5. Заключение

В заключение, потвърждавам, че представените материали отговарят напълно на изискванията на ЗРАСРБ, на Правилника за прилагане на ЗРАСРБ, на Правилника за условията и реда за придобиване на научни степени и за заемане на академични длъжности в Българската академия на науките, на Правилник за условията и реда за придобиване на научни степени и за заемане на академични длъжности в Института за български език „Проф. Любомир Андрейчин“ и убедено препоръчвам на уважаемата комисия да потвърди д-р Цветана Иванова Димитрова да заеме академичната длъжност доцент в направление 2.1. Филология, научна специалност Общо и сравнително езикознание за нуждите на Секцията за компютърна лингвистика в Института за български език „Проф. Л. Андрейчин“.

Дата 20 април 2023 г.

Изготвила становището:

(подпис)