

**АВТОРСКА СПРАВКА**  
за научните приноси  
на  
**Людмила Петрова Димитрова**

по конкурс за професор по научната специалност 01.01.12 Информатика (компютърна лингвистика – средства и системи за обработка на лингвистични знания)

**Забележка:** различните учебно-преподавателски дейности, които по-долу са прикачени към съответните текстове на **Правилника на ИМИ за приложение на ЗРАСРБ**, за удобство са дадени подредени и в отделен документ – Учебно-преподавателска дейност – **Учебна-резюме.pdf**

**По чл.29, т.3 от ЗРАСРБ, съотв. чл. 60 т.3 от ПЗРАСРБ - научни публикации:**

Общият брой на научните публикации е 64, от които 1 монографично изследване (вж. прил. **Публикации общо.pdf**)

За този конкурс са представени 29 научни публикации. В това множество няма публикации, представени за получаване на научната степен в 1977 г. Шест статии (1–6) са представени при получаване на академична длъжност „доцент“. Останалите 23 работи са публикувани след това – спазено е изискването на чл.2 т.6 от **Правилника на ИМИ за приложение на ЗРАСРБ**.

Представените за конкурса научни публикации (вж. прил. **Публикации за конкурса.pdf**) тематично попадат в следните направления: Компютърна лингвистика: (1) представяне и обработка на лингвистични знания и (2) разработка на цифрови многоезични корпуси, Средства и системи за обработка и управление на езикови ресурси, Модели на изследователски Е-инфраструктури. Допълнително е застъпено и направлението: Математическа лингвистика – логико-граматически формализми и компютърни системи за моделиране на резултати от теория на формалните езици.

**А. Математическа лингвистика – логико-граматически формализми и компютърни системи за моделиране на резултати от теория на формалните езици (3): 4, 5, 6**

Публикациите в тази група (с ясна теоретическа и практическата важност) са свързани с опитите за формално представяне на естествените езици, които продължават повече от 5 десетилетия след Ноам Чомски. Логическите граматики, създадени в периода 1978-1989, използват предикатната логика от първи ред за описание на естествено-езикови структури и явления, като осигуряват ефективен синтактичен и семантичен анализ и генериране на фрази при частична обработка или за ограничен фрагмент. Най-напред са използвани за моделиране на свободен словоред, координация, лява и дясна екстрапозиция и други езикови явления в английски език, по-късно – за испански и френски. В статията [5] задълбочено и последователно са представени четири вида логически граматики – *metamorphosis grammars*, *definite clause grammars*, *extraposition grammars*, *discontinuous grammars* и са разгледани възможностите им за описване на специфични особености на българския език. Статията [6] разглежда в детайли прекъснатите логически граматики. В работата за първи път е показано приложение на тези граматики за формалното описване на свободния словоред в българския език, предложен е и пример за описание на „управлението“ на относителността в граматичен смисъл, който илюстрира тясната връзка между лингвистичната теория и теорията на

формалните езици. Приносът е в използването за първи път на развити логико-граматически формализми за описание и моделиране на особено тежки за формално описание явления в български език (аспекти на синтаксиса), които имат пряко приложение към автоматизираната обработка на езика.

Близка по тематика, но с друга област на приложение е статията [4]. В първата ѝ част е дадена теоретичната база на описаната накратко във втората част експериментална програмна система. Доказани са теореми, свързващи определени типове формални езици с магазинни автомати. Програмната система е диалогова, предназначена е за моделиране на основни резултати от теорията на формалните езици и граматиките за нуждите на обучението. Приносът е научен (първата част на работата) и приложен (програмната система демонстрира различни еквивалентности при описание и генериране на формални езици).

## **Б. Компютърна лингвистика(12): 7, 10, 11, 12, 13, 14, 16, 21, 22, 24, 25, 27**

Публикациите в това направление са обособени в две подгрупи, които представят двата основни тематични раздела в областта: представяне и обработка на лингвистични знания и създаване на най-използваните в момента цифрови езикови ресурси – многоезичните корпуси.

### **Б1. Представяне и обработка на лингвистични знания (3): 10, 12, 13**

Материалите в тази група отразяват главно резултати от изследвания и разработки, изпълнени в рамките на проекта на ЕК *MULTEXT-East Multilingual Text Tools and Corpora for Central and Eastern European Languages* (МТЕ) и представят многоезични цифрови ресурси за 6 централни и източно-европейски езика: български, естонски, румънски, словенски, унгарски, чешки и английски като „опорен” език. Ресурсите включват средства за представяне и аотиране (в съответствие с международните стандарти) на лингвистични знания: морфосинтактични характеристики, лексикони, специфични (езиковозависими) ресурси, цифрови корпуси (едно- и многоезични, сравними, паралелни, аотирани и подравнени), предназначени за изследвания и приложения в областта на т. нар. езиковото инженерство. Разработките са свързани с научната и технологична готовност за включване на българския език в многоезични приложни системи (като напр. системи за машинен превод), в други проекти на Европейския съюз и пр.

Статията [10] представя средствата за описание и аотиране на лингвистични знания: хармонизирани лексически описания за шестте централно- и източно-европейски езици, разработени за МТЕ, с акцент на средствата за българския език. Лексическите описания MSD (**M**orpho**S**yntactical **D**escriptions) представят морфосинтактичните характеристики на всеки естествен език. Характеристиките са дефинирани като стойности на атрибути за всеки клас – определена част на речта съгласно граматиката на съответния език. Специфичните за всеки език явления, които нямат аналог в другите пет езика на проекта, са описани със същия формализъм, но са представени като отделен „специфичен“ езиков ресурс. По този начин се осигурява необходимата за автоматичния анализ – морфологичен, синтактичен или семантичен – информация за различните езици. В статията са описани и разработените лексикони (специализирани речници) и специфични (езиковозависими) ресурси за български език, необходими за сегментиране на аотирани езикови текстове. Лексиконите имат стандартна структура, съставени са от речникови статии, които съдържат заглавна дума (словоформа), основната форма на заглавната дума и съответното ѝ морфосинтактично описание (от множеството на хармонизираните МТЕ-описания). Българският МТЕ-лексикон се използва за обработка на българските текстове в МТЕ-корпуса. Като обем съдържа 55182 речникови статии (50810 словоформи, 4205 – съкращения и имена, 1044 числови данни).

Специфичните езикови ресурси за сегментиране на анотирани текстове са разработени като стандартни файлове, съдържащи за всеки език таблици с пунктуационните правила, най-употребяваните съкращения, съставните съкращения, клистиките и сложните съставни думи. Описаните ресурси се използват в системи за автоматично сегментиране на текст (т.нар. „сегментори“): от разпознаване на препинателните знаци до морфологичен анализ и преход „текст → речник“. Те са входни данни не само за „самообучаващи се“ софтуерни пакети, наречени „тагери“, но и в системи за „предсказване“ на възможното множество от тагове (определен тип характеристики) за непознати, предварително несрещани в текст думи, които липсват и в лексиконите. Разработеният формализъм дава възможност да се използват програмни средства, създадени в рамките на други европейски проекти (например MULTEXT *Multilingual Texts and Tools* за западно-европейските езици английски, италиански, испански, немски, финландски, френски и холандски,) за обработка и на български текстове, маркирани с SGML и снабдени със стандартните лексически описания (MSD). В статията е описано приложението на българските специфични МТЕ езикови ресурси за автоматично снемане на морфосинтактична многозначност.

В статия [12] са разгледани МТЕ морфосинтактичните спецификации на някои български глаголни форми (причастията) и са предложени нови, съответстващи на съвременната българска граматика.

В статията [13] са разгледани, от гледна точка на съпоставителната лингвистика, морфосинтактичните спецификации за български и за словашки език с цел хармонизация и анотиране на паралелни българо-словашки цифрови ресурси и за приложение в съпоставителните изследвания.

## **Б2. Разработка на цифрови многоезични корпуси (9): 7, 11, 14, 16, 21, 22, 24, 25, 27**

### **Разработка на първите български цифрови анотирани корпуси (3): 7, 11, 25**

Статията [7] (доклад, представен в 1998 г. на най-представителния форум по компютърна лингвистика COLING-ACL'98) отразява създаването на всички цифрови езикови ресурси, вкл. корпусите, в рамките на проекта MULTEXT-East на ЕС (<http://nl.ijs.si/ME>). Ресурсите са разработени в стандартизиран формат, хармонизиращ специфичните характеристики на всеки от шестте езика на проекта, със стандартно маркиране и анотиране съгласно препоръките на TEI (Text Encoding Initiative) Working Group. Паралелният МТЕ корпус продължава да служи като **модел на многоезичен подравнен анотиран ресурс** при разработване на нови корпуси за други естествени езици (вж. материалите **по чл.2 т.6 от Правилника на ИМИ за приложение на ЗРАСРБ**, стр. 8, 9 на настоящата Авторска справка).

В [11] и [25] са представени в развитие цифрови ресурси с български език, като накратко е описан и разработения за проекта MULTEXT-EAST първи български електронен корпус, включващ **паралелен и подравнен корпус** МТЕ-1984.Bg (романът “1984” на Джордж Оруел) и **сравним корпус от 2 части** (МТЕ-fiction.bg и МТЕ-news.bg). Корпусът МТЕ-1984.Bg съдържа около 300 хил. думи, маркирани с SGML в CES-формат. Част от МТЕ-1984.Bg е **анотиран на ниво словоформа корпус** с обем 87235 словоупотреби: към всяка словоформа от текста на българския превод на „1984“ е присъединена лингвистична информация, включваща нейната основна форма (лема) и съответното ѝ морфосинтактично описание (MSD). Анотираният корпус е получен след автоматично снемане на синтактична многозначност и пост-редакция от автора за отстраняване на грешките. МТЕ-Fiction.Bg е сравним корпус, с обем 97251 словоупотреби, ръчно анотиран от автора на ниво параграф, като собствените съществителни имена са маркирани за „тип“: лице, място и др. Анотираният корпус са

ценен ресурс за „самообучаващи се“ софтуерни пакети за автоматична обработка на естествен език.

В рамките на проекта TELRI (<http://www.ids-mannheim.de/telri/>) беше издаден двутомен CD-ROM „East meets West: A Compendium of Multilingual Language Resources“, който се разпространява за изследователски цели. Вторият том на това издание съдържа описание на резултатите на проекта MULTEXT-East, вкл. примерни образци на корпуса в HTML-формат.

### **Разработка на цифрови многоезични паралелни и подравнени корпуси (6): 14, 16, 21, 22, 24, 27**

#### ***Първи българо-полски корпус (2): 14, 16***

[16] представя поредна версия на първия българо-полски/полско-български цифров корпус, разработен от авторите в рамките на междуакадемичното сътрудничество между ИМИ-БАН и Института по славистика на ПАН. В работата са обсъдени проблемите на разработка на паралелни корпуси. Двyezичният корпус съдържа художествена литература (белетристика, публицистика) и специална лексика (текстове на документи на Европейската Комисия и на Европейския Съюз). Корпусът е съставен от две части: паралелен (в обем над 3 млн. словоупотреби) и сравним (текстове на български и полски език, сравними по отношение на техния обем). Част от паралелните текстове са анотирани (художествена литература – на ниво „изречение“; специална лексика, с включен трети език – английски, на ниво „параграф“).

В статията [14] е направено сравнение между две системи за анотиране на текстове (на ниво словоформа) на два славянски езика: МТЕ за български език и IPI-PAN за полски език.

#### ***Първи българо-словашки подравнен корпус (1): 27***

В статията [27] са представени два двyezични корпуси, съставени от паралелни текстове на три славянски езика – български, полски и словашки. Това са първите българо-полски/полско-български и българо-словашки/словашко-български корпуси, разработени от тримата автори. Първият паралелен българо-словашки/словашко-български корпус е разработен от Людмила Димитрова и Радован Гарабик в рамките на междуакадемичното сътрудничество между ИМИ-БАН и Института по лингвистика „Людовит Щур“ на САН. Корпусът, с обем 1 млн. словоупотреби, съдържа два вида текстове: оригинални текстове на български език или на словашки език с преводите им, съответно, на словашки или на български, и текстове, преводи на български и словашки от трети език. Част от паралелните текстове са подравнени на ниво „изречение“ и образуват първия подравнен българо-словашки/словашко-български корпус, който съдържа около 500 хил. словоупотреби. Подравненият корпус е достъпен в Интернет: с помощта на заявка на български или на словашки език се задава словоформа – ключ, с който се извършва търсенето в корпуса. В резултат потребителят получава конкорданс, с който може да проследи значението на интересуващата го дума в един широк контекст („изречение↔изречение“), а не само в рамката „дума↔дума“. В статията са описани подробно приложенията на двyezичните корпуси в съпоставителните изследвания – за съпоставяване на славянски езици: един аналитичен (български) с два синтетични (полски и словашки).

#### ***Първи българо-полски-литовски паралелен корпус (3): 21, 22, 24***

Три статии, [21], [22] и [24], са посветени на първия българо-полски-литовски корпус (с обем около 3 млн. словоупотреби). Той съдържа две части: паралелен (около 1 млн. словоупотреби) и сравним корпус. Част от паралелния корпус съдържа текстове (белетристика), подравнени на ниво „изречение“ и текстове (специална лексика), подравнени на ниво „параграф“, които образуват подравнен триезичен корпус. Сравнимият корпус е съставен от текстове,

публикувани от електронни медии в Интернет, които са сравними не по обем, а по съдържание: те описват едно и също събитие. Като опорен език към трите езика е присъединен английски език. Подравненият триезичен корпус дава възможност за съпоставяне на езици от две различни езикови групи: славянската (български и полски) и балтийската (литовски). Триезичният корпус има широко приложение в съпоставителните изследвания.

Статията [21] е достъпна на Интернет-страницата на международната организация Association for Computational Linguistics (ACL): <http://www.aclweb.org/anthology/W/W09/W09-4001.pdf>

## **В. Средства и системи за обработка и управление на езикови ресурси (11): 1, 2, 3, 8, 9, 15, 17, 18, 19, 20, 29**

В това направление са обособени три подгрупи:

### **В1. Автоматична обработка на български език (3): 1, 2, 3**

В тази група са събрани описанията на пионерните разработки за автоматична обработка на български език с доказан научен и научно-приложен принос в областта.

В [1] е представена първата система за автоматично сегментиране на българските словоформи без използване на лексемен речник (речник на българските основи) и без каквито и да е ограничения, наложени върху входния текст. Системата реализира описания в [2] конструиран от авторите първи алгоритъм за автоматичното сегментиране (определяне на основа и окончание) на словоформите в текста. Алгоритъмът е комбинация от евристични съображения и точно определени множества от езикови елементи (множествата на българските окончания), той е разработен въз основа на анализа на 700000 български словоформи, които като обем представят съдържанието на няколко тълковни речника и могат да се образуват от 70-те хиляди единици на Обратния речник на българския език. Системата съдържа и съответните обслужващи програми за създаване, поддържане и обновяване на базовата лингвистична информация. Разработката е използвана съществено при създаването на пакет приложни програми „Автоматично сегментиране на български словоформи“ (вж. материалите по чл.2 т.6 от Правилника на ИМИ за приложение на ЗРАСРБ, стр. 8, 9 на настоящата Авторска справка).

Статия [3] представя две версии на програмна система за автоматичен анализ на български текст. За първи път за български език е реализиран алгоритъм за автоматично снемане омонимията на окончанията (автоматично причисляване на дадена словоформа към определен граматичен клас). Системата, работеща в режим на диалог на български език, дава възможност за изследване вида на лингвистичната информация, която може да се извлече от текста в резултат на този вид автоматичен анализ. Системата реализира автоматичен преход „текст → речник“, включващ сегментиране на български текст и автоматично построяване на различни видове речници и речникови архиви: конкорданси на даден текст или даден автор с добавяне на натрупаната в процеса на обработка статистическа информация, честотни речници, комбинирани речници. По зададена от потребителя заявка, системата реализира търсене и автоматично извличане на словоформи и словосъчетания от даден български текст, с контекст или не, с цел изучаване поведението в текста на дадената лексическа единица, синтактична конструкция или определен граматичен клас. Системата е разработена в две версии (зависещи от наличните по това време технологични платформи): едната версия е реализирана на APPLE-PASCAL за персонален компютър APPLE II, а другата е реализирана на PL/1 в DOS за EC1040.

## **B2. Разработка на лексически бази данни (3 ): 8, 17, 20**

### ***Лексическа база данни за български език (1): 8***

Проектът PL96-1142 CONCEDE Consortium for Central European Dictionary Encoding на програмата INCO-COPERNICUS има за цел да хармонизира методологията, средствата и ресурсите за създаване на обща лексическа база данни (LDB) за шест европейски езика – български, чешки, естонски, унгарски, румънски и словенски. Разработените ресурси са преносими и многократно използвани, тъй като използват стандарта на Text Encoding Initiative Dictionary Working Group и са валидирани чрез прилагане на ISO8879SGML. CONCEDE LDB е предназначена за интегрални изследвания и приложения в системи за езиково инженерство, машинен превод и обучение. Схемата от лексикографски спецификации съгласно международните стандарти за кодиране на електронни речници е разработена за проекта CONCEDE с цел използване в многоезични приложни системи на ЕС.

В [8] са отразени резултатите от изследванията за българския език в рамките на проекта, а именно разработка на лексикографските спецификации за българския език съгласно международните стандарти за кодиране на електронни речници, подбора на заглавните думи за едноезичната лексическа база данни за български език (като част от многоезичната база данни CONCEDE), комплектоването и кодирането на речниковите статии за тези думи. CONCEDE LDB за българския език съдържа 2700 речникови единици, избрани по определена методика, според изискванията на проекта, базирана е на Българския тълковен речник и използва лексикографските спецификации, разработени в рамките на проекта.

Разработката на LDBs за българския език е определена като едно от най-важните научно-приложни постижения на ИМИ-БАН през 2000 г.

### ***Разработка на многоезични бази данни (2): 17, 20***

Приносът на [17] се състои в разработка на модел на двуезична лексическа база данни (LDB). Работата отразява проектирането и разработването на първата двуезична LDB за поддръжане на българо-полски/полско-български онлайн речници. Описан е и синтактичен анализатор за трансфер на данните от LBD в релационния модел на речника.

За реализирането на синтактичния анализатор са избрани съвременни решения и технологии, използвани за разработка на web-базирани приложения, осигуряващи висока скорост и гъвкавост на управлението на LBD.

В статията [20] са представени резултатите от работата върху многоезична терминологична база данни на термините от областта на корпусната лингвистика. Софтуерната среда за представяне на терминологична база данни е базирана на PHP и Python, което позволява разширяване на обема на данните, както по отношение на броя на езиците, така и по отношение на съдържанието.

## **B3. Двуезични онлайн българо-полски/полско-български речници (5): 9, 15, 18, 19, 29**

Приносът на първите три статии в тази група е в поставянето за дискусия на проблемите, свързани с класификаторите на заглавната дума в речниковата статия на цифровите двуезични речници. В [9] подробно и методично, с цел използване в многоезични системи за машинен превод, са обсъдени проблемите, свързани с хармонизацията на класификаторите в речниковата статия на цифровите многоезични речници. В [15] е описано представянето на българския глагол в цифровите българо-полски/полско-български речници. Предложени са нови класификатори (разграничаващи форми и значения) за описване на специфични

характеристики и езикови явления в българския език. Дискусията за стандартизирано представяне на класификаторите продължава в [18]. Обсъдени са примери от текущата експериментална версия на българо-полския онлайн речник.

Приносът на [19] и [20], представящи първия българо-полски онлайн речник, е научно-приложен. Статията [19] описва разработването на двуезичния онлайн речник. Web-приложението за представяне на речника (реализирано с два основни програмни модула: „администратор“ и „за потребителя“) е базирано на LDB, чието проектиране и разработване е описано детайлно в [17]. При определени условия (1-1 преводно съответствие на заглавните думи) програмите реализират и автоматично създаване на речникова статия за полско-български онлайн речник. Експерименталният вариант на речника работи оптимално на Internet Explorer 6.0+ (Windows), Firefox 2.0.1+ (Windows, Linux). Избраната технология предоставя възможности за първоначално инициране на релационния модел с данни от LDB, които да служат за последващо модифициране, разширяване и визуализиране от web-базираното приложение на речника. Текущото състояние на разработката, с разширена и обновена LDB, е представено в [29].

### **Г. Модели на изследователски Е-инфраструктури (3): 23, 26, 28**

Материалите в това направление отразяват работата по европейския проект 211938 *MONDILEX Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and Their Digital Resources* на 7-ма Рамкова програма на ЕК.

Статията [23] описва целите и задачите на инициатива за създаване на модел на изследователска Е-инфраструктура за езикови ресурси, разработени за шест славянски езика: български, полски, руски, словашки, словенски и украински. В работата са представени и постиженията на участващите в инициативата институции в областта на езиковите ресурси и технологии.

В [28] са представени накратко най-важните резултати от изпълнението на проекта MONDILEX, вкл. неговия научен и социален импакт. Специално място е отделено на възможностите, които предлагат Grid-базираните технологични платформи за обработка на стандартизирани многоезични ресурси.

Приносът на монографичното изследване [26] е в предложения концептуален модел на Grid-базирана Европейска изследователска Е-инфраструктура, поддържаща цифрови езикови ресурси за славянска лексикография. Моделът обхваща всички съвременни езикови технологии и видове цифрови езикови ресурси, които трябва да бъдат включени в такава изследователска Е-инфраструктура: корпуси (едно- и многоезични, паралелни и съпоставими, анотирани и подравнени), речници (едно- и двуезични, електронни и онлайн), бази данни, лексикони, тезауруси, онтологии. Разгледани са стандартите TEI за представяне на цифрови езикови ресурси, създадени с цел тяхното многократно използване в съвременни многоезични системи. Като технологична платформа за поддръжка на изследователската Е-инфраструктура е предложена една модерна среда – Knowledge Grid, приложение на Grid-технологията за обмен и обработка на многоезични ресурси. Предназначението на Концептуалния модел е да интензифицира съвместните научни изследвания в областта на цифровата лексикография, осигурявайки възможност за разпространение на иновациите независимо от географското положение.

Разработката на Концептуалния модел е определена като едно от най-важните научно-приложни постижения на ИМИ-БАН през 2010 г. (Вж. отчет на БАН за 2010: [http://www.bas.bg/fce/001/0079/files/OTCET\\_BAN\\_2010.pdf](http://www.bas.bg/fce/001/0079/files/OTCET_BAN_2010.pdf))

**По чл.2 т.6 от Правилника на ИМИ за приложение на ЗРАСРБ:  
- трудове на други учени, публикувани в авторитетни издания, които съществено  
използват резултати на кандидата:**

Публикации [1] и [2] от списъка на публикациите за конкурса са използвани при разработката на Пакет приложни програми за автоматично сегментиране на български текст:

Kiril Simov, Galia Angelova (1985). A **Program Package for Automatical Segmentation of Bilgarian Wordforms**. In Proceedings of the First Conference of Program Designers, July 11-12, 1985, Budapest, Ed. By A. Iványi, 19-23.

Публикации [16], [21], [22], [24] от списъка на публикациите за конкурса са използвани при разработката на триезичния корпус с български, полски и руски език:

M. Duškin, J. Satoła-Staškowiak. **The Bulgarian-Polish-Russian Parallel Corpus**. In: International Journal Cognitive Studies/Études Cognitives. Vol. 11, SOW, Warsaw, 2011. 241-254. ISSN 2080-7147.

Публикация [13] от списъка на публикациите за конкурса е цитирана в статиите, включени в сборниците от доклади, представени на авторитетни конференции:

1. Christian Chiacaros and Tomaz Erjavec (2011). OWL/DL formalization of the MULTTEXT-East morphosyntactic specifications. In: *Proceedings of the 5th Linguistic Annotation Workshop (LAW-V), held in conjunction with the ACL-HLT 2011*, June 2011, Portland, Oregon, USA, p. 11-20. <http://aclweb.org/anthology-new/W/W11/W11-0402.pdf>
2. Christian Chiacaros (2010). Grounding an Ontology of Linguistic Annotations in the Data Category Registry. In: *Proceedings of the Language Resource and Language Technology Standards – state of the art, emerging needs, and future developments, 18th of May, 2010 within LREC'2010*, Malta. <http://www.lrec-conf.org/proceedings/lrec2010/workshops/W4.pdf>
3. Alexandr Rosen (2010). Morphological tags in parallel corpora. In: Čermák, František, Klégr, Aleš, & Corness, Patrick (eds.). 2010. In: *Proceedings of the InterCorp: Exploring a Multilingual Corpus*. Praha: Nakladatelství Lidové noviny.
4. T. Erjavec (2010). MULTTEXT-East Version 4: Multilingual Morphosyntactic Specifications, Lexicons and Corpora. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC'10, 19-21 May 2010*, Malta. 2544-2547. Publ. by ELRA, Paris. ISBN 2-9517408-6-7 [http://www.lrec-conf.org/proceedings/lrec2010/pdf/138\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/138_Paper.pdf)
5. Darja Fišer and Beñot Sagot. (2008) Combining Multiple Resources to Build Reliable Wordnets. In *Proceedings of Text Speech and Dialogue Conference, TSD'08*, Brno. 61-68. **Springer-Verlag Berlin Heidelberg 2008**, ISSN 0302-9743/ISBN-10 3-540-87390-2 **Springer Berlin Heidelberg NewYork**/ISBN-13 978-3-540-87390-7 Springer Berlin Heidelberg NewYork <http://www.springerlink.com/content/j20t5p166654/front-matter.pdf>

Публикация [7] от списъка на публикациите за конкурса е цитирана в статии, публикувани в авторитетни издания:

1. T. Erjavec (2011). **MULTTEXT-East: Morphosyntactic resources for many languages**. In: *International Journal Language Resources and Evaluation*, 2011. 27 pp. (in print)
2. Homa Baradaran Hashemi, Azaden Shakery, Hesham Faili (2010). **Creating a Persian-English Comparable Corpus**. In Proceedings of the International Conference of the Cross-Language Evaluation Forum, CLEF 2010, Padua, Italy, September 2010. **Multilingual and Multimodal Information Access Evaluation, Lecture Notes in Computer Science, vol. 6360, Springer 2010**, 27-39,
3. Daniel Jurafsky and James H. Martin. (2010) **Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition**. 1024 pages, Publisher: Prentice Hall; 2 edition (May 26, 2008), ISBN-10: 0131873210, ISBN-13: 978-0131873216
4. TamasVaradi. (2007) **NP Modification Structures in Parallel Corpora**. In: Incorporating Corpora: The Linguist and the Translator, 2007 - books.google.com Pp. 168-186 Type: eBook Document; English. Publisher: Clevedon ; Buffalo : Multilingual Matters, ©2008. ISBN: 9781853599873 1853599875 <http://scholar.google.bg/scholar?hl=en&lr=&cites=13996482698111050937>
5. Julia Trushkina. (2007) **Development of a Multilingual Parallel Corpus and a Part-of-Speech Tagger for Afrikaans**. In: **IFIP International Federation for Information Processing, Volume 228, Intelligent Information Processing III**, eds. Z. Shi, Shimohara K., Feng D., (Boston: Springer), pp. 453-462. 2007. ISBN 978-0-378-44641-7-47
6. Julia Trushkina. (2006) **The North-West University Bible corpus: A multilingual parallel corpus for South African languages**. In: **Language Matters: Studies in the Languages of Africa, Volume 37, Issue 2, 2006, pages 227 – 245**. Publishing House Routledge. ISSN: 1753-5395 (electronic), ISSN:1022-8195 (paper) <http://www.informaworld.com/10.1080/10228190608566262>
7. Behrang QasemiZadeh, Saeed Rahimi. (2006) **Persian in MULTTEXT-East Framework**. In: Advances in Natural Language Processing, Proceedings of the 5th International Conference on NLP, FinTAL 2006, Turku, Finland, August 23-25, 2006. **Lecture Notes in Computer Science, vol. 4139, Springer 2006**, 541-551, DOI: 10.1007/11816508\_54. ISBN 3-540-37334-9. Also at: <http://www.springerlink.com/content/g23370g0w02m00r3>
8. Daniel Varga, Peter Halacsy, Andras Kornai, Viktor Nagy, Laszlo Nemeth, Viktor Tron. (2005) **Parallel Corpora for Medium Density Languages**. Proceedings of the International Conference RANLP-2005. Borovec, Bulgaria, 21-23 September 2005. 590-596. ISBN 954-91743-3-6. (A version of this paper appeared in: Nicolov et al (Eds): **Current Issues in Linguistics Theory, Volume 292, Recent Advances in Natural Language Processing IV**. Selected papers from RANLP 2005. 247-258. John Benjamins Publ. Company Amsterdam, Philadelphia, 2007, ISBN 978 90 272 4807 7.)
9. András Kuba, András Hócza, and János Csirik. (2004) **POS Tagging of Hungarian with Combined Statistical and Rule-Based Methods**. In Petr Sojka, Ivan Kopeček, and Karel Pala, editors, Text, Speech and Dialogue, Proceedings of the 7th International Conference, TSD 2004, Brno, Czech Republic and LNAL. **Lecture Notes in Computer Science, vol. 3206, Springer, 2004**, 113-120, DOI: 10.1007/b100511 <http://springerlink.metapress.com/openurl.asp?genre=article&issn=0302-9743&volume=3206&page=113>



10. András Kuba, László Felföldi, András Kocsor. (2005) **POS tagger combinations on Hungarian text**. In Proceedings of the Second International Joint Conference on Natural Language Processing, IJCNLP, Jeju Island, Korea, October 11-13, 2005, Companion Volume including Posters/Demos and tutorial abstracts, 191-196, <http://acl.ldc.upenn.edu/I105/I05-2033.pdf>, also in **Lecture Notes in Computer Science**, vol. 3651, Springer 2005, ISBN 3-540-29172-5
11. F. Mihelič, J. Gros, S. Dobrišek, J. Žibert, N. Pavešić. (2003) **Spoken Language Resources at LUKS of the University of Ljubljana**. In **International Journal of Speech Technology**, Vol. 6, Num. 3, 221-232, 2003, DOI:10.1023/A:1023462002932. **Kluwer Academic Publishers**. Manufactured in The Netherlands.

Публикация [7] от списъка на публикациите за конкурса е цитирана в следните дисертации:

1. Camelia Ignat. (2009) Improving Statistical Alignment and Translation Using Highly Multilingual Corpora. June 2009, **Université de Strasbourg, France**. 215 pp.
2. L. Felföldi (2008). Classifier Combination Systems and their Application in Human Language Technology. **University of Szeged, Doctoral School in Mathematics and Computer Science, Ph.D. Programme in Informatics**, May 2008, 118 pages
3. Victoria Bobicev. (2007) METODE ȘI ALGORITMI STATISTICI DE PROCESARE A TEXTELOR - Statistica - **PhD Thesis 2007. Universitatea Tehnică a Moldovei** Cu titlu de manuscris CZU: 004.912 [http://www.cnaa.acad.md/files/theses/2007/6478/viorica\\_bobicev-thesis.pdf](http://www.cnaa.acad.md/files/theses/2007/6478/viorica_bobicev-thesis.pdf).
4. Anna Feldman. (2006) Portable language technology: a resource-light approach to morpho-syntactic tagging. **PhD Thesis, 2006, Ohio State University**, <http://etd.ohiolink.edu/> <http://etd.ohiolink.edu/send-pdf.cgi/Feldman%20Anna.pdf?osu1153344391>
5. Markus Dickinson. (2005) Error detection and correction in annotated corpora. **Ph.D. thesis, 2005, Ohio State University**. <http://etd.ohiolink.edu/> [http://etd.ohiolink.edu/view.cgi?acc\\_num=osu1123788552](http://etd.ohiolink.edu/view.cgi?acc_num=osu1123788552)

### По чл.3 от Правилника на ИМИ за приложение на ЗРАСРБ:

#### - ръководство и участие в международни и национални научноизследователски проекти:

1. Координатор (ръководител на Консорциума) на проект от 7 Рамкова програма на Европейската комисия: EC FP7 GA 211983 project *MONDILEX Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and Their Digital Resources* (2008 – 2010), EC FP7 Capacities – Research Infrastructures
2. Участник (представител на ИМИ) в EU Research Infrastructure *CLARIN Common Language Resources and Technology Infrastructure* (2007 – 2011), EC FP7 Capacities – Research Infrastructures
3. Участник (представител на ИМИ) в БГ-КЛАРИН *Национална интердисциплинарна изследователска Е-инфраструктура за интегриране и развитие на електронните ресурси за български език като част от Европейския CLARIN* (2010 – 2015), Пътна карта за Национални изследователски инфраструктури
4. Ръководител на проект „Семантика и съпоставителна лингвистика, ориентирани към разработване на двуезичен електронен речник“ (Semantics and Contrastive Linguistics with a Focus on a Bilingual Electronic Dictionary) с Полската академия на науките, Институт по славистика (2006 – 2008, 2009 – )
5. Ръководител на проект „Електронни корпуси – съпоставително изследване с цел проектиране на българо-словашки електронни езикови ресурси“ (Electronic Corpora – Contrastive Study with Focus on Design of Bulgarian-Slovak Digital Language Resources) със Словашката академия на науките, Институт по лингвистика „Людовит Щур“ (2008 – )
6. Водещ изследовател в проект PL96-1142 project CONCEDE *Consortium for Central European Dictionary Encoding* (1998 – 2000), EU 5<sup>th</sup> FP INCO-COPERNICUS Programme
7. Водещ изследовател в проект EU COP 106 project *MULTEXT-EAST Multilingual Text Tools and Corpora for Central and Eastern European Languages* (1995 – 1997), EU COPERNICUS Programme
8. Участник в проект EU COP 200 *ELSNET GOES EAST* (1995 – 1996), EU COPERNICUS Programme
9. Участник в ELSNET: *European Network of Excellence in Language and Speech* (1993 – )
10. Участник в проект „Математически проблеми на информатиката“ (Mathematical Problems of Informatics) с Унгарската академия на науките, Изследователски институт по автоматизация и изчислителна техника (1992 – 2009)
11. Участник в проект „Изследвания на модели и програмни средства за езикови технологии“ (Изследвания на модели и програмни средства за езикови технологии) с Руската академия на науките, Институт по проблеми на предаването на информацията (2008)
12. Участник в проект „Изследвания и разработки в областта на българо-украински електронни езикови ресурси“ (Изследвания и разработки в области българско-украински електронни езикови ресурси) с Националната академия на науките на Украйна, Лингво-информационен фонд на НАНУ (2008 – 2009)

13. Ръководител на национален изследователски проект „Нови информационни и езикови технологии в машинния превод” (1995 – 1998), Договор И-526/95 с МОНТ
14. Участник в проект „Алгоритми, методи и средства за обработка на естествени езици”, ИМИ-БАН, (1996 – )
15. Участник в проект „Създаване на нови методи и инструментални средства на изкуствения интелект – математическа лингвистика и обработка на естествени езици”, ИМИ-БАН, (1988 – 1993)
16. Участник в национален изследователски проект „Съвременни методи на изкуствения интелект в обработката на естествените езици” (1991 – 1994), Договор И-14/91 с МОН
17. Участник в национален изследователски проект „Интерфейс човек-машина на естествен език. Функционални езикови процесори” (1988 – 1991), Договор № 27 с МНВО
18. Участник в проект „Математическа кибернетика и бази данни” (Mathematical Cybernetics and Databases) с Унгарската академия на науките, Изследователски институт по автоматизация и изчислителна техника, (1985)
19. Водещ участник в проект „Методологични проблеми на граматиката, семантиката и лексикографията” с Централен институт по езикознание на АН на ГДР и ИМИ-БАН (1981 – 1985)
20. Водещ участник в проект „Семантични критерии и структури” с Централен институт по езикознание на АН на ГДР и ИМИ-БАН (1978 – 1980)

**- участие в програмни и организационни комитети на научни мероприятия (от последните 5 години):**

**Програмни комитети на международни научни мероприятия:**

The Sixth International Conference SLOVKO'2011 *NLP, Multilinguality*, Modra-Bratislava, Slovakia, 20-22 October 2011, член на програмния комитет

MONDILEX Fifth Open International Workshop „Research Infrastructure for Digital Lexicography” in conjunction with International Conference Information Society – IS 2009, Ljubljana, Slovenia, 14 – 15 October 2009, председател на програмния комитет

MONDILEX Fourth Open International Workshop „Representing Semantics in Digital Lexicography” in conjunction with International Scientific Conference COGNITIVE and CONTRASTIVE STUDIES, Warsaw, Poland, 29 June – 1 July 2009, председател на програмния комитет

MONDILEX Third Open International Workshop „Metalanguage and Encoding Scheme Design for Digital Lexicography”, Bratislava, Slovak Republic, 15 – 16 April 2009, председател на програмния комитет

MONDILEX Second Open International Workshop „Organization and Development of Digital Lexical Resources”, Kiev, Ukraine, 2 – 4 February 2009, председател на програмния комитет

MONDILEX First Open International Workshop „Lexicographic tools and techniques”, Moscow, Russia, 3 – 4 October 2008, председател на програмния комитет

Jubilee International Conference on Mathematical and Computational Linguistics „30 years Department of Mathematical Linguistics”, 6 July, 2007, Sofia, Bulgaria, член на програмния комитет

**- членство в авторитетни творчески и/или професионални организации в съответната научна област:**

1. Association for Computational Linguistics
2. EWM *European Woman in Mathematics*

**- участия с доклади в международни и национални научни форуми (от последните 6 години):**

**2011:**

1. L. Dimitrova, R. Garabik (2011). Bulgarian–Slovak Parallel Corpus. International Conference SLOVKO'2011 *NLP, Multilinguality*, 20-22 October 2011, Modra, Slovakia (предстои)
2. L. Dimitrova, V. Koseska–Toszewa, D. Roszko, R. Roszko (2011). Bulgarian-Polish-Lithuanian Corpus – Recent Progress and Application. International Conference SLOVKO'2011 *NLP, Multilinguality*, 20-22 October 2011, Modra, Slovakia (предстои)

3. L. Dimitrova, R. Dutsova, R. Panova (2011). Survey on Current State of Bulgarian-Polish Online Dictionary. International Workshop "Language Technologies for Digital Humanities and Cultural Heritage" within International Conference RANLP'2011 *Recent Advances in Natural Language Processing*, 12-16 September 2011, Hissar, Bulgaria
4. L. Dimitrova, R. Dutsova, R. Panova (2011). Information Technologies for the Preservation of Language Heritage. International Conference. International Conference DiPP 2011 Digital Preservation and Presentation of Cultural and Scientific Heritage, 11-14 September 2011, Veliko Tarnovo, Bulgaria
5. L. Dimitrova (2011). Language Resources – a Part of World Cultural Heritage. Information Technologies for the Preservation of Language Heritage. International Conference. International Conference DiPP 2011 Digital Preservation and Presentation of Cultural and Scientific Heritage, 11-14 September 2011, Veliko Tarnovo, Bulgaria
6. L. Dimitrova (2011). Parallel Corpora with Bulgarian texts. International Workshop on Language Corpora in Bulgaria – Diversity and Application, 21 May 2011, Sofia, Bulgaria
- 2010:**
7. L. Dimitrova (2010). Project MONDILEX. IncoNet East European and Central Asian Countries Project Coordinators Board Meeting, 25-26 November 2010, Bonn, Germany
- 2009:**
8. L. Dimitrova (2009). Bulgarian Digital Resources in a Multilingual Context. International Conference of the EC Directorate General for Translation "*Influence of European Integration on Slavic Languages – Trends in Translation and Multilingual Communications*", 2 December 2009, Bratislava, Slovakia, available at [http://ec.europa.eu/slovensko/documents/archive/konferencia\\_02122009/prof.\\_dimitrova.pdf](http://ec.europa.eu/slovensko/documents/archive/konferencia_02122009/prof._dimitrova.pdf)
9. L. Dimitrova (2009). From Electronic Corpora to Online Dictionaries (on the example of Bulgarian Language Resources). International Conference *SLOVKO'2009*, 25-27 November 2009, Bratislava, Slovakia
10. L. Dimitrova, V. Koseska-Toszewa, D. Roszko, R. Roszko (2009). Bulgarian-Polish-Lithuanian Corpus. International Conference "*Bulgarian Studies in Poland – Positions and Prognoses*", 6th November 2009, Krakow, Poland
11. L. Dimitrova, V. Koseska-Toszewa, D. Roszko, R. Roszko (2009). Bulgarian-Polish-Lithuanian Corpus – Problems of Development and Annotation. MONDILEX Fifth Open Workshop "Research Infrastructure for Digital Lexicography" within International Conference *Information Society 2009*, 16-17 October 2009, Ljubljana, Slovenia
12. L. Dimitrova, V. Koseska-Toszewa, D. Roszko, R. Roszko (2009). Bulgarian-Polish-Lithuanian Corpus – Current Development. International Workshop "Multilingual resources, technologies and evaluation for Central and Eastern European languages" within International Conference *RANPL'2009*, 17 September 2009, Borovec, Bulgaria, available at <http://www.aclweb.org/anthology/W/W09/W09-4001.pdf>
13. L. Dimitrova, V. Koseska-Toszewa, R. Dutsova, R. Panova (2009). Bulgarian-Polish Online Dictionary – Design and Development. MONDILEX Fourth Open Workshop "Representing Semantics in Digital Lexicography" within International Scientific Conference *COGNITIVE and CONTRASTIVE STUDIES*, 29 June -1 July 2009, Warsaw, Poland
14. M. Šimková, R. Garabík, L. Dimitrova (2009). Design of a multilingual terminology database prototype. MONDILEX Fourth Open Workshop "Representing Semantics in Digital Lexicography" within International Scientific Conference *COGNITIVE and CONTRASTIVE STUDIES*, 29 June -1 July 2009, Warsaw, Poland
15. L. Dimitrova, R. Panova, R. Dutsova (2009). Lexical Database of the Experimental Bulgarian-Polish online Dictionary. MONDILEX Third Open Workshop "Metalanguage and Encoding Scheme Design for Digital Lexicography". 15-16 April 2009, Bratislava, Slovakia
16. L. Dimitrova, V. Koseska-Toszewa and J. Satoła-Staškowiak (2009). Towards a Unification of the Classifiers in Dictionary Entry. MONDILEX Third Open Workshop "Metalanguage and Encoding Scheme Design for Digital Lexicography". 15-16 April 2009, Bratislava, Slovakia
17. L. Dimitrova, R. Garabík and D. Majchráková (2009). Comparing Bulgarian and Slovak Multext-East morphology tagset. MONDILEX Second Open Workshop "Organization and Development of Digital Lexical Resources", 2-4 February 2009, Kiev, Ukraine
18. L. Dimitrova, P. Rashkov (2009). A New Version for Bulgarian MULTEXT-East Morphosyntactic Specifications for Some Verbal Forms. MONDILEX Second Open Workshop "Organization and Development of Digital Lexical Resources", 2-4 February 2009, Kiev, Ukraine

19. L. Dimitrova, V. Koseska–Toszewa, I. Derzhanski, R. Roszko (2009). Annotation of Parallel Corpora. MONDILEX Second Open Workshop “Organization and Development of Digital Lexical Resources”, 2-4 February 2009, Kiev, Ukraine

#### **2008**

20. L. Dimitrova, V. Koseska–Toszewa (2008). The significance of entry classifiers in digital dictionaries. MONDILEX Open Workshop “Lexicographic Tools and Techniques”, 3–4 October, 2008, Moscow, Russia

21. L. Dimitrova, R. Pavlov (2008). On Compatibility of Slavic Language Resources. MONDILEX Open Workshop “Lexicographic Tools and Techniques”, 3–4 October, 2008, Moscow, Russia

#### **2007**

22. L. Dimitrova (2007). Bulgarian Digital Resources as a Base for Automatic Disambiguation. International Conference “Cognitive Studies: Semantics and Contrastive linguistics with a focus on multilingual electronic dictionaries”, 10 – 11 May 2007, Warsaw, Poland

23. L. Dimitrova, I. Derzhanski (2007). Representation of Mode of Action in the Bilingual Digital Dictionary (with). International Conference “Cognitive Studies: Semantics and Contrastive linguistics with a focus on multilingual electronic dictionaries”, 10 – 11 May 2007, Warsaw, Poland

24. L. Dimitrova, V. Koseska–Toszewa (2007). Digital Dictionaries – Problems and Features. Jubilee International Conference „Mathematical and Computational Linguistics”, 6 July 2007, Sofia, in conjunction with „New Trends in Mathematics and Informatics” - Jubilee International Conference 60 years IMI, BAS. July 2007, Sofia, Bulgaria.

25. L. Dimitrova, R. Pavlov (2007). Digital Bulgarian Language Resources - Specifications, Corpora, Lexica. Jubilee International Conference „Mathematical and Computational Linguistics”, 6 July 2007, Sofia, in conjunction with „New Trends in Mathematics and Informatics” - Jubilee International Conference 60 years IMI, BAS. July 2007, Sofia, Bulgaria.

#### **2006:**

26. L. Dimitrova, I. Derzhanski and E. Sendova (2006). Electronic Lexicography and Its Applications: The Bulgarian Experience. International Scientific Conference MegaLing 2006: *Horizons of Applied Linguistics and Linguistic Technologies*. 20 – 27 September 2006, Parthenit, Ukraine

#### **- участия в редколегии на научни издания:**

##### ***Редколегия на международно списание (2010 – )***

Member of the Editorial Board (with Violetta Koseska (Editor-in-Chief, Poland), Wiesław Banyś (Poland), Denis Pierre Apothélos (France), Björn Hansen (Germany), Roman Laskowski (Poland)) of the International Journal Cognitive Studies/Études Cognitives. Published by SOW Publ. House, Warsaw. ISSN 2080-7147. IF ERIH for Social sciences and Humanities. (справка CS/EC.pdf) [http://www.ispan.waw.pl/content/view/224/99/lang.pl\\_PL.ISO8859-2/](http://www.ispan.waw.pl/content/view/224/99/lang.pl_PL.ISO8859-2/)

##### ***Редактор на научни издания:***

Violetta Koseska and Antoni Mazurkiewicz. Time Flow and Tenses. (Ludmila Dimitrova, Zbigniew Grén, Editors), SOW Publ. House, Warsaw, Poland. 223 pages. 2010. ISBN 978-83-89191-94-6.

Volodymyr Shyrokov, Ludmila Dimitrova (Editors 2009), Organisation and Development of Digital Lexical Resources. Proceedings of the MONDILEX Second Open Workshop, Kiev, Ukraine, 2–4 February 2009. Dovira Publ. House, Kiev, Ukraine. ISBN 978-966-507-252-2. [www.mondilex.org/pro2.pdf](http://www.mondilex.org/pro2.pdf)

Violetta Koseska, Ludmila Dimitrova, Roman Roszko (Editors 2009). Representing Semantics in Digital Lexicography. Proceedings of the MONDILEX Fourth Open International Workshop, Warsaw, Poland, 29 June–1 July 2009. SOW, Warsaw, 2009, 76-88. ISBN 978-83-89191-87-8. [www.mondilex.org/pro3.pdf](http://www.mondilex.org/pro3.pdf)

Leonid Iomdin, Ludmila Dimitrova (Editors 2008). Lexicographic Tools and Techniques. Proceedings of the MONDILEX First Open Workshop, Moscow, 3-4 October 2008. IITP Publ. House, Moscow, Russian Federation. ISBN 978-5-9900813-6-9. [www.mondilex.org/pro1.pdf](http://www.mondilex.org/pro1.pdf)

L. Dimitrova & R. Pavlov (Editors 2007). Mathematical and Computational Linguistics: Proceedings of the Jubilee International Conference dedicated to the 30th anniversary of the Mathematical Linguistics Dept., 6 July 2007, Sofia, Bulgaria, in conjunction with New Trends in Mathematics and Informatics - Jubilee International Conference dedicated to the 60th anniversary of the IMI - BAS , 6-8 July 2007, Sofia, Bulgaria. 2007. 136 pages. ISBN 978-954-8986-28-1.

#### **- аудиторни занятия във висши училища – лекции и семинари:**

### **Четени курсове в Софийски университет, ФМИ, специалност: информатика**

Дискретна математика, 120 часа годишно, упражнения, 1978 – 1990

Математически основи на информатиката, 60 часа годишно, упражнения, 1988 – 1989

Математическа и компютърна лингвистика, 30 часа годишно лекции и 60 часа годишно упражнения, 1991 – 1994

### **Четени курсове в Софийски университет, ФКНФ Магистърска програма**

Математически и логически основи на компютърната лингвистика, 30 часа годишно лекции и 30 часа годишно упражнения, 1995 – 1997

### **Четени курсове в НБУ, бакалавърски факултет, специалност: информатика**

Автомати, езици, изчисление. 30 часа годишно лекции и 60 часа годишно упражнения, 1996 – 2000 (Приложена справка за четените лекции в НБУ)

### **Четени курсове в ИМИ - ВТУ “Св. Св. Кирил и Методий” Магистърска програма “Езикови и мултимедийни технологии” (Приложена справка за часовете по програмата):**

Логика, език, информация. 30 часа лекции. 2002, 2003, 2004, 2005, 2006, 2007

Компютърна лексикография. 45 часа практикум. 2002, 2003, 2004, 2005, 2006, 2007

Професионални комуникации. 15 часа лекции и 15 часа упражнения. 2002, 2003, 2004, 2005

Програмиране в Интернет. 45 часа практикум. 2002, 2003, 2004, 2005

Дискретна математика и елементи на математическата логика. 45 часа лекции и 30 часа упражнения. 2004, 2005, 2006

Компютърни системи и архитектури. 30 часа лекции и 15 часа упражнения. 2004, 2005, 2006;

Бази данни. 30 часа лекции и 15 часа упражнения. 2004, 2005, 2006

Програмиране в Интернет и професионални комуникации. 30 часа лекции и 30 часа упражнения. 2006, 2007.

### **- разработване на лекционни курсове (създадени и прочетени за първи път в България курсове във ВУ):**

Компютърна лексикография – ИМИ-ВТУ Магистърска програма “Езикови и мултимедийни технологии”, 2001

Логика, език, информация – ИМИ-ВТУ Магистърска програма “Езикови и мултимедийни технологии”, 2001

Автомати, езици, изчисление – НБУ бакалавърски факултет, 1996.

### **- дейности, свързани с научното развитие на докторанти, дипломанти и студенти:**

#### *Докторанти – ръководство*

№	Име	Тип	Защитил
1	Ралица Дуцова Ralitsa Dutsova	Задочна	Текуща

#### *Магистранти – ръководство*

№	Име Name	Теза Title of Thesis /Supervised Master Theses: M.Sc. in Informatics/	Година на защита Year of defense
1	Ирина Ненова	Автоматично сегментиране на българските словоформи	1981
2	Борислав Иванов	Автоматично получаване на лингвистична информация от линейния контекст /за български текстове	1982
3	Невяна Исусова	Система за автоматично получаване на лингвистична информация	1985
4	Мая Енчева	Експериментална програмна система за моделиране работата на крайни автомати и дясно-линейни граматика	1990
5	Валентина Пътова	Експериментална програмна система за моделиране на безконтекстни граматика в диалогов режим	1990
6	Мариела Тодорова	Експериментална програмна система за моделиране на магазинни автомати и безконтекстни граматика в диалогов режим	1990
7	Евелина Божанкова	Експериментална програмна система за компютърно моделиране на машина на Тюринг	1992.
8	Детелина Данова	Уеб-базирано приложение за представяне на образователни програми	2007
9	Румяна Панова	Бази данни – лексическа и релационна – на българо-полски онлайн речник	2009
10	Ралица Дуцова	Уеб-базирано приложение за представяне на Българо-полски речник	2009

### Приложни разработки – създаване на информационните продукти:

1. MTE Language Resources: Version 4: <http://nl.ijs.si/ME/V4/doc/index.html#sec-crp>
2. Slovak-Bulgarian Corpus Linguistics Terminology Data Base: <https://data.juls.savba.sk/mltd/>;  
<http://data.juls.savba.sk/mltd/>
3. Bulgarian-Slovak corpus on the web: <http://korpus.sk:8090/>
4. Bulgarian-Polish corpus,
5. Bulgarian-Polish-Lithuanian corpus,
6. Bulgarian-Polish Electronic Dictionary,
7. Bulgarian-Polish Online Dictionary

### Участие в експертна дейност, рецензиране:

1. Участник в Научно жури по чл. 4 ал. 2 от ЗРАСРБ и чл. 2 ал. 8 от ППЗРАСРБ за конкурс за заемане на академична длъжност „професор“
2. Рецензент на:
  - Договори с МОНТ- 3 броя
  - Публична рецензия на програмни продукти и системи: Експертна оценка на прототип на електронна платформа, разработена по договор LLP-TOI-BGI-08-LEO-05-00437 “Training Safety and Security in Tourism”, (БСУ, 2009 г.) (1 рецензия, 2009 г.)
  - Рецензент на 4 магистърски тези по специалност 01.01.12 Информатика, представени за защита в рамките на Магистърската програма „Мултимедийни и езикови технологии“:

№	Име	Теза	Година на защита
1	Станислава Славова	Интерактивни среди за обучение по електронна търговия	2004
2	Владимир Георгиев	<i>Дигитални библиотеки с мултимедийно съдържание</i>	2006
3	Васил Бадев	Архитектура и изграждане на цифрова библиотека за комплексни мултимедийни обекти от българското фолклорно наследство	2007
4	Детелин Лучев	Описание на семантиката на значими области от българското фолклорно наследство чрез технологии за обработка на знания	2007