



30
години

**Секция за компютърно
моделиране на
българския език**

<https://dcl.bas.bg/news/>

20
години

**Секция по
компютърна
лингвистика**



Секция по компютърно моделиране



**Основател на Секцията по
компютърно моделиране на
българския език през 1994 г.:
проф. д.ф.н. Йордан Пенчев**



Секция по компютърна лингвистика



**Ръководител на Секцията от
2001 г. до днес: проф. д-р Светла
Коева**







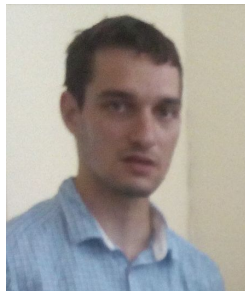
Членовете на Секцията по компютърна лингвистика



доц. д-р
Цветана Димитрова



доц. д-р
Емил Дойчев



гл. ас. д-р
Йордан Кралев



гл. ас. д-р
Христина Кукова



гл. ас. д-р
Светлозара Лесева



гл. ас. д-р
Михаела Москова



гл. ас. д-р
Валентина Стефанова



гл. ас. д-р
Ивелина Стоянова



Емил Стоянов



гл. ас. д-р
Мария Тодорова



Димитър Христов



Мартин Ялъмов



Сътрудници на Секцията по компютърна лингвистика през 2024 г.



проф. д.ф.н. Елена Иванова
проф. д.ф.н. Максим Стаменов
проф. д-р Мила Димитрова-Вълчанова
проф. д-р Илияна Кръпова
проф. д-р Йовка Тишева
проф. д-р Тинко Тинчев
доц. д-р Иван Держански
доц. д-р Марина Джонова
гл. ас. д-р Олена Сирук
гл. ас. д-р Росица Декова
д-р Георги Джумайов
д-р Виктория Петрова
Димитър Христов



Сътрудници на Секцията по компютърна лингвистика през годините



проф. д.м.н. Стоян Михов
проф. д-р Радка Влахова
доц. д-р Атанас Атанасов
доц. д-р Анета Динева
доц. д-р Петя Несторова
доц. д-р Екатерина Търпоманова
гл. ас. д-р Георги Чолаков
д-р Димитър Георгиев
д-р Иванка Петрова
Катя Алахверджиева
Анелия Белогай
Светлана Горанова
Кристина Душанова

Албена Копринарова
Николай Раданов
Петя Плачкова
програмист Кузман Белев
програмист Александър Ботов
програмист Гечо Гечев
програмист Ангел Генов
програмист Борислав Митев
програмист Никола Обрешков
програмист Борислав Ризов
програмист Борислав Станков
програмист Павел Ташев
програмист Димитър Христов

Докторанти



Докторанти на Секцията



Анета **Динева** (докторант на Й. Пенчев)

Светла **Коева** (докторант на Й. Пенчев)

Цветомира **Венкова** (докторант на Й. Пенчев)

Петя **Плачкова** (докторант на Й. Пенчев)

Петя **Несторова** (докторант на Й. Пенчев)

Иванка **Петрова** (докторант на Й. Пенчев)

Светозара **Халачева** (докторант на Й. Пенчев)

Милен **Русков** (докторант на Й. Пенчев)

Диана **Милиева** (докторант на Й. Пенчев)

Галина **Петрова** (докторант на Й. Пенчев)

Кристина Калпакчиева (докторант на Т. Тинчев)

Борислав **Ризов** (докторант на Т. Тинчев)

Ангел **Генов** (докторант на А. Дичев)





Докторанти на Секцията



Христина **Кукова** (докторант на Р. Влахова)
Никола **Обрешков** (докторант на С. Стоянов)
Николай **Желязков** (докторант на Й. Тишева)
Ивелина **Стоянова** (докторант на С. Коева)
Светлозара **Лесева** (докторант на С. Коева)
Мария А. **Тодорова** (докторант на С. Коева)
Георги **Илиев** (докторант на С. Коева)
Лъчезар **Джаков** (докторант на С. Коева)
Тодор **Лазаров** (докторант на С. Коева)
Георги **Джумайов** (докторант на С. Коева)
Виктория **Петрова** (докторант на С. Коева)
Димитър **Христов** (докторант на С. Коева)



Научни проекти



Проекти с външно финансиране



1994–1996: **LaTeSlav**

2002–2004: **O CoRrect**

1912–2014:



1997–2002: **TELRI**[®]

2007–2009: **MTrans**

2014–2015:

LASI

1997–1998:



2008–2013:



2015–2022:



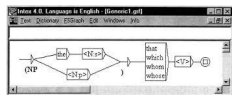
1999: **ELAN**

2008–2011: **LexIt**

2021–2022:



1999–2001:



2009–2011:



2023:



1999–2006: **Daskal**

2014–2015: **MediaTalk**

2023:





Двустранни международни проекти



С **Института по славистика** към ПАН от 1994 година досега, с **Университета Париж-IV: Пари-Сорбон** (1994–2000), с **Университета Франш Конте в Безансон** (2005–2006), с **Института за чешки език** към ЧАН (2001–2006).



С **Центъра за изкуствен интелект** към Румънската академия:
Обогатени бази от знания за български и румънски (2012 – 2017)
Обогатяване на лингвистични ресурси с деривационно свързани несвободни фрази (2018 – 2020)

С **Института за компютърна лингвистика „А. Замполи“** към Националния съвет за научни изследвания на Италия:
Оценка на уменията за четене и разбиране в начална училищна възраст в България и Италия (2023–2025)



Оценка на
уменията
за четене



Проекти, финансирани от CEF на ЕК



Проекти за усъвършенстване на системите за автоматичен превод, финансирани от Механизма за свързване на Европа на Европейската комисия, Програма на Сектор „Телекомуникации“ (2017 до 2022)

За системата „Преводач“ за Българското председателство на Съвета на ЕС: *Автоматичният превод на Механизма за свързване на Европа за Председателството на Съвета на Европа.*



За системите на Платформата за автоматичен превод на Механизма за свързване на Европа (CEF.AT).



Многоезикови ресурси за CEF.AT от областта на законодателството (MARCELL) и колекция от многоезикови ресурси за CEF.AT (CURLICAT).



Големи европейски проекти



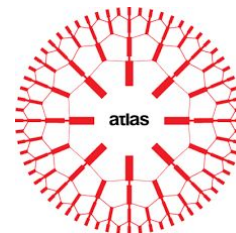
Проекти, финансирани по Четвъртата рамкова програма и Програмата за подкрепа на политиките в областта на информационните и комуникационните технологии на ЕС

BalkaNet – Design and Development of a Multilingual
Balkan WordNet (2001–2004)



Central and South-east european Resources
(CESAR) (2011–2013)

Applied Technology for Language-Aided CMS (ATLAS)
(2010–2013)





Големи европейски проекти



Проекти, финансирани по Програма *Хоризонт 2020* и целево от Европейския съюз

European LEXicographic InfraStructure
(ELEXIS) (2019–2022)



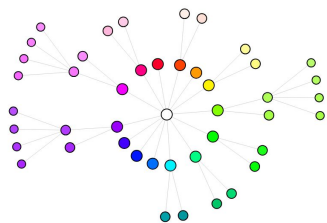
Multilingual Image Corpus (MIC21)
(2021–2022)



Теоретични проекти, финансирани от ФНИ



2017 – 2020



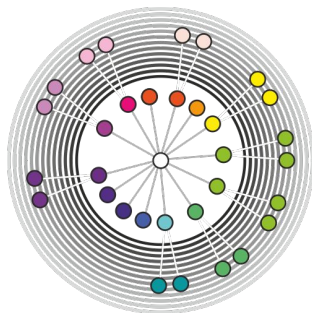
Семантична мрежа с широк спектър от семантични релации

2020 – 2023



Онтология на ситуациите за състояние – лингвистично моделиране.
Съпоставително изследване за български и руски

2021 – 2023



Обогатяване на Семантичната мрежа
Уърднет с концептуални фреймове

2024 – 2026



Онтология на предикатите за действие – лингвистично моделиране с фокус върху българския език



Дългосрочни проекти за езикови ресурси



Дългогодишни проекти, посветени на разработване на езикови ресурси

Глаголна семантика: семантико-синтактичен речник на българския език (2005–2010)

**Българската лексикалносемантична мрежа
БулНет (2005–2016)**



Български национален корпус (2009–2016)

**Семантични ресурси и програми
за обработка на езика: БулНет и ФреймНет (2017–2026)**



Езикови ресурси



Българският уърднет (БулНет)

<https://dcl.bas.bg/bulnet/>



БулНет 3.0 БулНет & PWN en ro Login

Synset: bg - n: звезда:3

дефиниция: небесно тяло, представляващо масивно газово (плазмено) кълбо с определена маса, размер, плътност, светимост и химически състав, което излъчва собствена светлина вследствие на термоядрени реакции във вътрешността

лтерал: звезда:3

част на речта: n ili: eng-30-09444100-n +0 -0
семантичен клас: noun.object

хипероним: bg - n: небесно тяло:1; космически обект:1

хипоним: bg - n: Слънце:1; слънце:4

хипоним: bg - n: слънце:3

хипоним: bg - n: двойна звезда:1; бинарна звезда:1

хипоним: bg - n: неподвижна звезда:1

хипоним: bg - n: гигант:3; звезда гигант:1

хипоним: bg - n: пътеводна звезда:1

хипоним: bg - n: множествена звезда:1

Synset: en - n: star:7

дефиниция: (astronomy) a celestial body of hot gases that radiates energy derived from thermonuclear reactions in the interior

лтерал: star:7

част на речта: n ili: eng-30-09444100-n +0 -0
семантичен клас: noun.object

хипероним: en - n: celestial body:1; heavenly body:1

хипоним: en - n: binary star:1; binary:4; double star:1

хипоним: en - n: fixed star:1

хипоним: en - n: giant star:1; giant:4

хипоним: en - n: lodestar:2; loadstar:2

хипоним: en - n: multiple star:1

хипоним: en - n: neutron star:1

хипоним: en - n: nova:1

хипоним: en - n: red dwarf:1; red dwarf star:1

Synset: sl - n: zvezda:1; zvezda:3

дефиниция:

лтерал: zvezda:3

лтерал: zvezda:1

част на речта: n ili: eng-30-09444100-n +0 -0
семантичен клас: noun.object

хипероним: sl - n: astr:1; nebesno telo:2

хипоним: sl - n: dvojna:1; dvojni:4; dvojna:1

хипоним: sl - n: severna:1; severna:4; severna:1

хипоним: sl - n: nevtro:1; nevtro:4; nevtro:1

хипоним: sl - n: nova:1; nova:4; nova:1

хипоним: sl - n: rdeča:1; rdeče pritlikavke:1

хипоним: sl - n: K2:1; m:1; rdeča orjakinja:1; rdeča vel:1

Synset: ro - n: stea:2

дефиниция: Corp ceresc cu lumină proprie, format dintr-o masă de gaze aflată la o temperatură foarte ridicată

лтерал: stea:2

част на речта: n ili: eng-30-09444100-n +0 -0
семантичен клас:

хипероним: ro - n: astru:1; corp_ceresc:1

хипоним: ro - n: stea_binară:1

хипоним: ro - n: stea_gigant:1

хипоним: ro - n: stea_călușitoare:2; steaua_polară:1

хипоним: ro - n: stea_multiplă:1

хипоним: ro - n: stea_neutron:1

хипоним: ro - n: novă:1

хипоним: ro - n: pitică_roșie:1

хипоним: ro - n: gigant_roșu:1

хипоним: ro - n: soare:2



Български национален корпус

<https://dcl.bas.bg/bulnc/>



Над **1,2 млрд. думи** на български
Паралелни корпуси на над 40 езика с
общ обем от **5,4 млрд. думи**

Автоматична лингвистична анотация
на българската част, основна анотация
и на други езици

Система за търсене по езикови и
метаезикови критерии

Система за автоматична многослойна
анотация, достъпна като свободна
услуга

Няма защита | search.dcl.bas.bg

 **BNC** Български национален корпус
Bulgarian National Corpus

Връзки

Асистент Подкорпуси

Намерени 90308 Ляв Десен

1.	И тръгна покрай гигантската змийска кожа .	+
2.	Пътеката се губеше в гъстата утринна мъгла .	+
3.	Той е древната човешка памет .	+
4.	Докосна дългата дървена дръжка .	+
5.	Единствената бледа луна бе изчезнала.	+
6.	...пада Можеш да си вземеш нов Пикъл В този пад е единствената оригинална версия на eXistenZ Цялата система която струва 3...	+
7.	... в пълна сила застъпена тук, на борда на световно известната Бойна Звезда Галактика .	+
8.	Седяха в малката тясна стаичка зад бара.	+
9.	Това била отходната канализационна тръба .	+



Анотирани корпуси



Корпус от съвременни текстове на български с обем от **174 697 думи**, в който на всяка словоформа еднозначно е приписана граматична информация от Българския граматичен речник и е определена основната ѝ форма.



Корпус, отразяващ състоянието на съвременната българска лексика, с обем от близо **100 000 лексикални единици**, в който на всички думи е приписано лексикално значение от Българския уърднет.

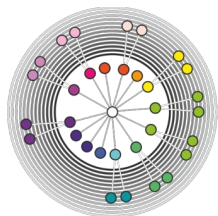


Паралелен двуезиков корпус от текстове на български и английски език със съотнесени прости изречения в състава на сложните и анотация на вида на синтактичната връзка между двойките прости изречения (над **360 000 думи** за двата езика).



Българският фреймнет

<https://dcl.bas.bg/bulframe/>



Семантично представяне на глаголите чрез система от абстрактни концептуални фреймове, съотнесени със семантичните фреймове на Фреймнет и семантичната мрежа Уърднет.

Пълно и непротиворечиво описание на:

езиково специфичната задължителна съчетаемост между глаголите и техните обкръжения;

семантичните ограничения, наложени върху задължителните обкръжения:

синтактичната реализация на задължителните обкръжения (синтактична категория и граматична функция).

Word	
бъбря	
Characteristic	Value
Frame element	included
Element name	Speaker
Status	core
Semantic type	Sentient
Syntactic obligatoriness	prodrop
Syntactic category	included
NP	included
Frame element	included
Element name	Addressee
Status	core
Semantic type	Sentient
Syntactic obligatoriness	no
Syntactic category	included
PP	included



Многоезиков корпус с изображения

<https://dcl.bas.bg/MIC-21/>



**MULTILINGUAL
IMAGE CORPUS
2021**

Голяма колекция от свободни изображения с очертани маски на над **200 000 обекта** и класификация в съответствие с *Онтология на визуалните обекти*.

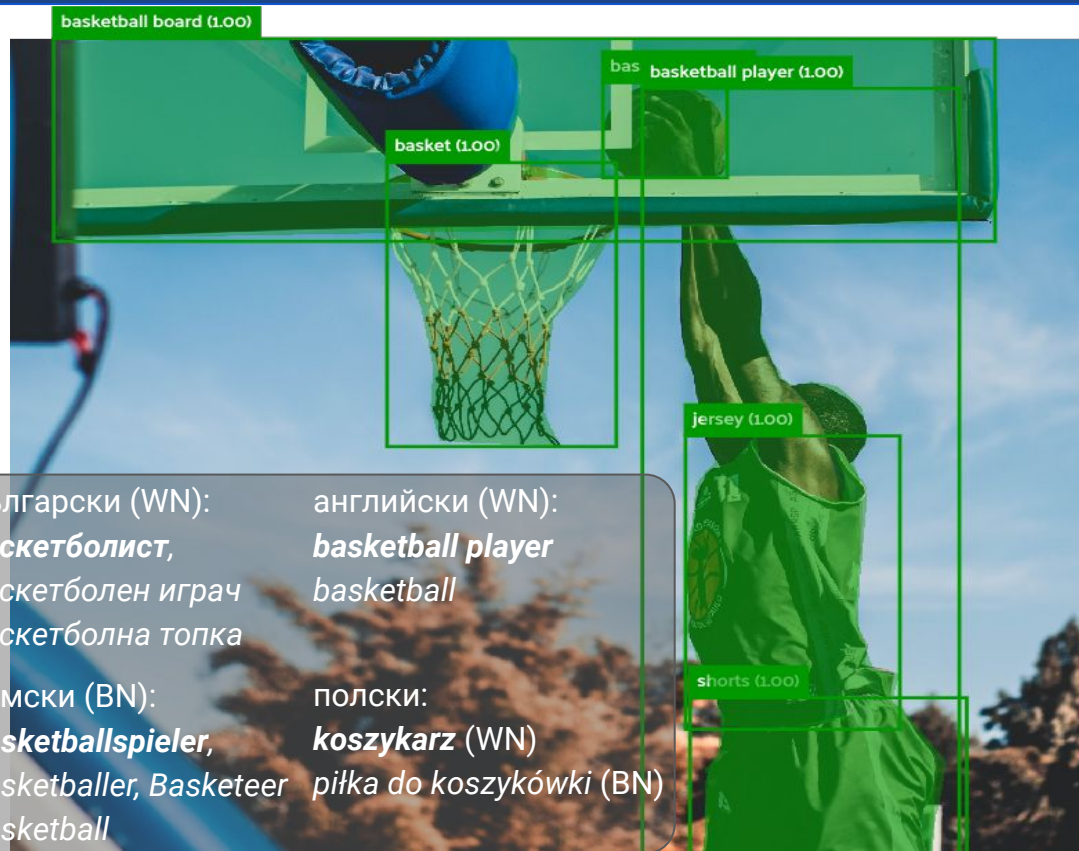
Класовете на обектите в изображенията са съотнесени с многоезиково описание, основаващо се на Уърднет.

български (WN):
баскетболист,
баскетболен играч
баскетболна топка

немски (BN):
Basketballspieler,
Basketballer, Basketeer
Basketball

английски (WN):
basketball player
basketball

полски:
koszykarz (WN)
piłka do koszykówki (BN)



Научни сборници и книги



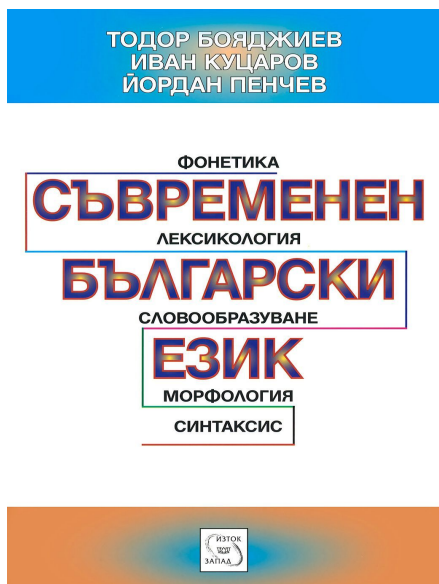
Трудове на членовете на Секцията



1998



1998



1999



2004

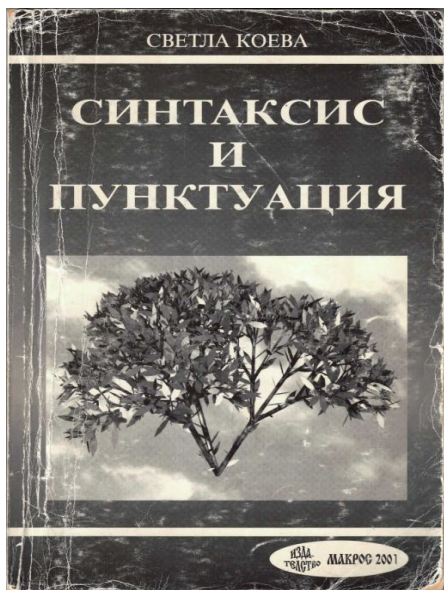




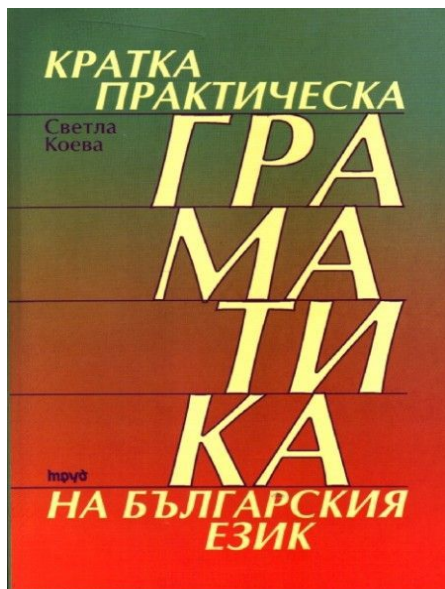
Трудове на членовете на Секцията



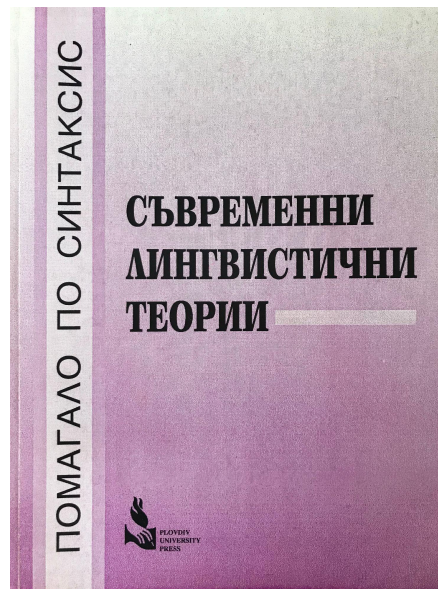
1999



2001



2001



2005

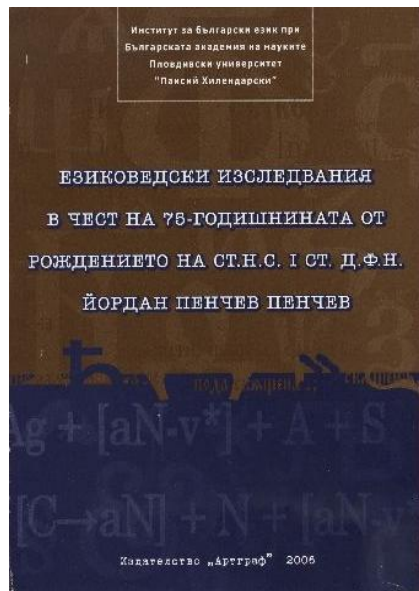




Сборници на Секцията



2006



2008



2010



2014

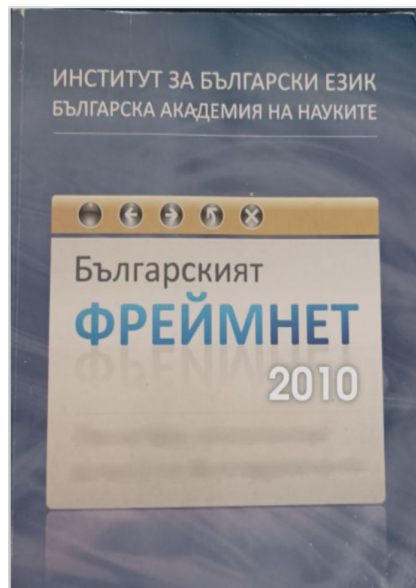




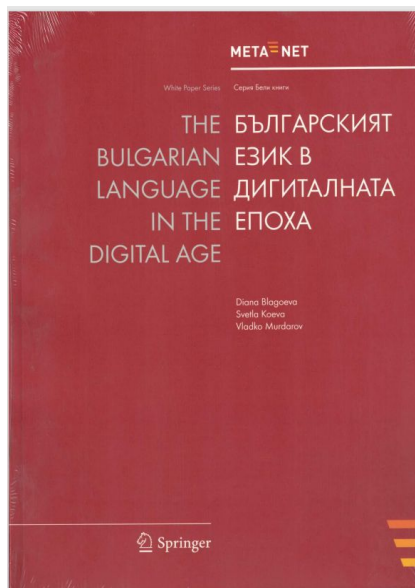
Трудове на членовете на Секцията



2010



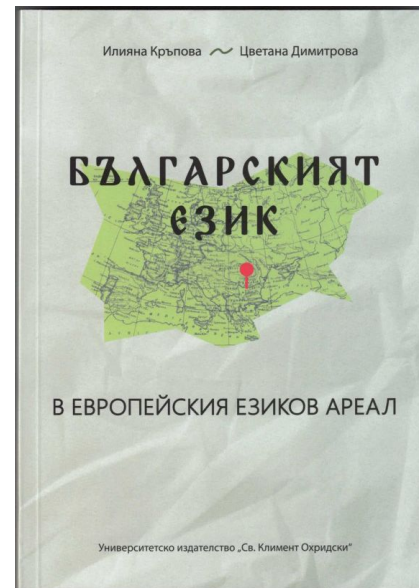
2012



2023



2024

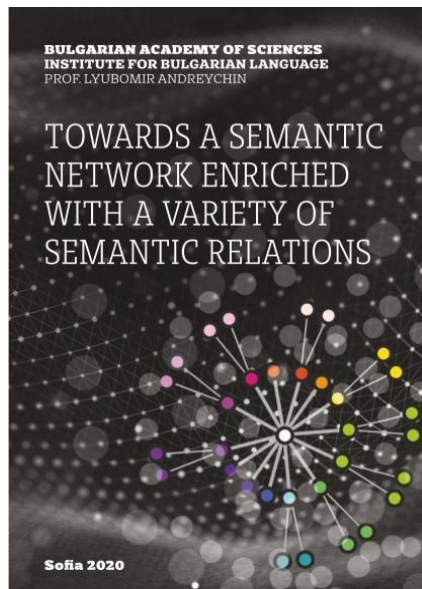




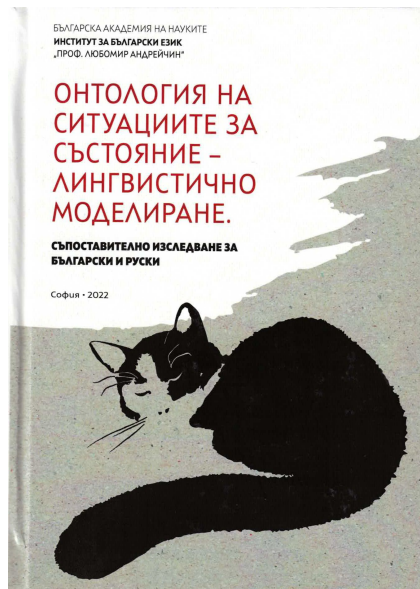
Сборници на Секцията



2020



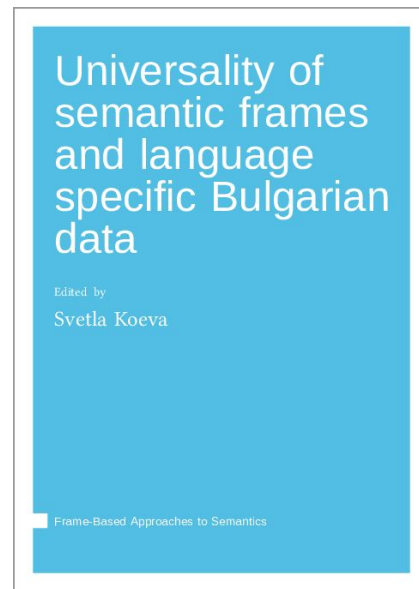
2022



2024

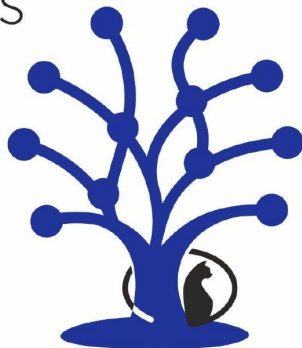


2024



ISSN: xxx-xxx-xxx

Computational
Linguistics
In
Bulgaria



Editor-in-chief:
Prof. Svetla Koeva

ISSN 3033-1382 (печатно издание)
<https://jclib.dcl.bas.bg>

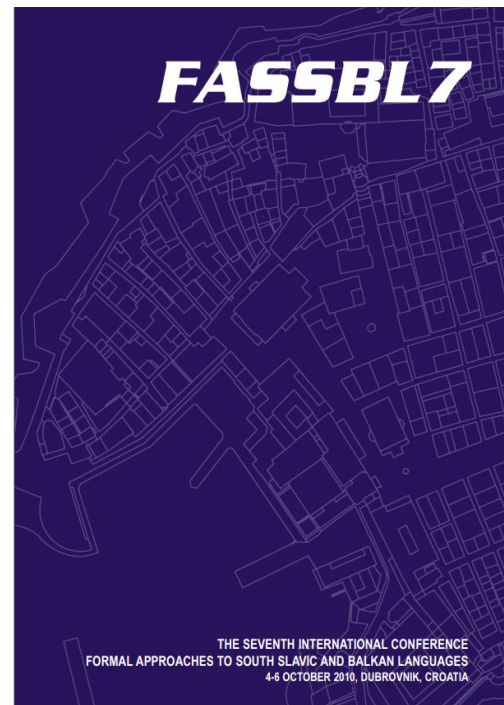
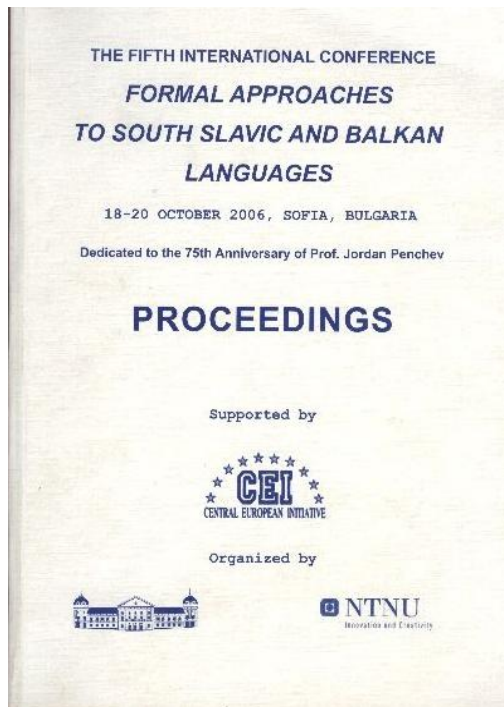
Department of Computational Linguistics
Institute for Bulgarian Language *Prof. Lyubomir Andreychin*
Bulgarian Academy of Sciences



Организовани конференции, научни сесии,
семинари



Формални подходи към южнославянските и балканските езици (FASSBL) 2006 – 2012





51-вата годишна конференция на Асоциацията по компютърна лингвистика



The Association for
Computational
Linguistics

ACL 2013
4-9 August
Sofia, Bulgaria

WELCOME TO ACL 2013!

For the first time, the annual meeting of the Association for Computational Linguistics (ACL) takes place in Bulgaria. ACL 2013 will be held in Sofia, Bulgaria's capital, August 4-9, 2013. As in previous years, the program of the conference includes a poster session, tutorials, workshops and demonstrations in addition to the main conference.

- About ACL 2013
- Welcome message
- Overview
- Committees
- Venue
- Important dates
- Programs



51-вата годишна конференция на Асоциацията по компютърна лингвистика



Юбилейна научна сесия
„Съвременни тенденции в
езиковедските изследвания“,
посветена на 85 години от
рождението на проф. д.ф.н.
Йордан Пенчев
в рамките на
Международната
конференция Паисиеви
четения 2016





Международната конференция *Компютърната
лингвистика в България* CLIB 2014 – 2024

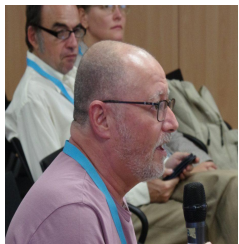
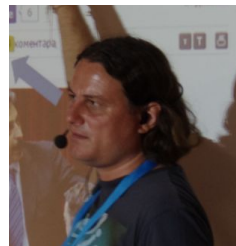
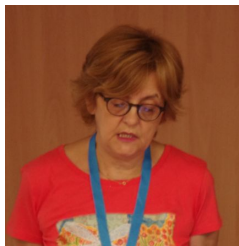


Computational Linguistics in Bulgaria

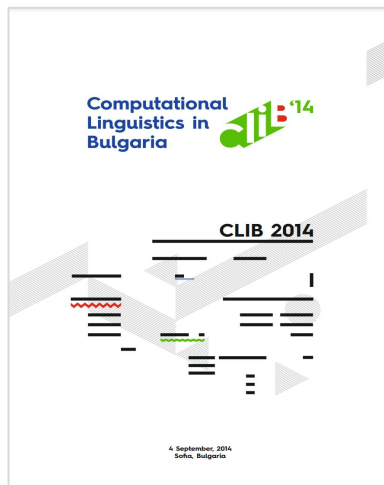




Поканени лектори



Сборниците с
доклади се
индексират в
Scopus,
Web of Science,
ACL Anthology.



PROCEEDINGS



PROCEEDINGS



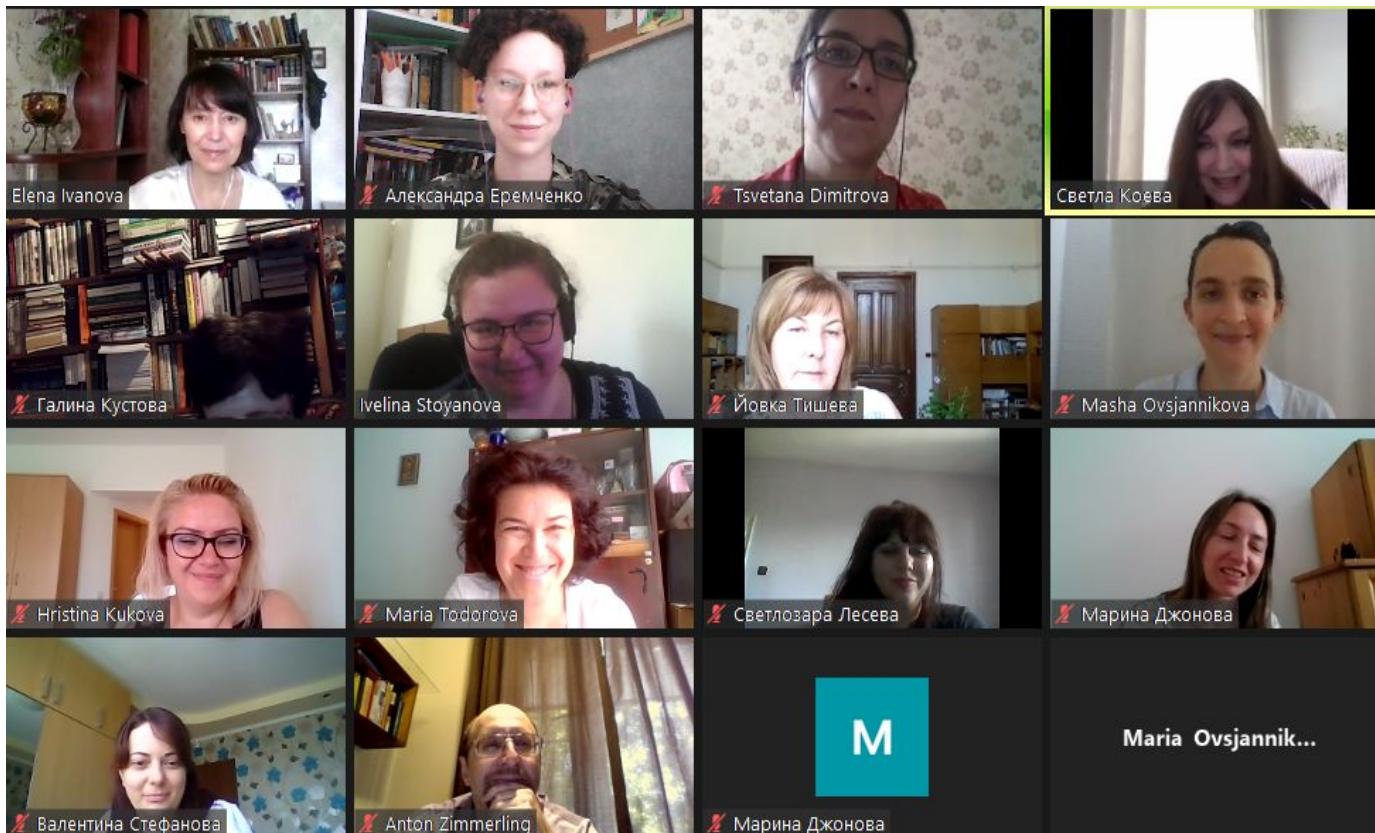
PROCEEDINGS

Department of Computational Linguistics
Institute for Bulgarian Language
BULGARIAN ACADEMY OF SCIENCES





Научна сесия на Осми форум „Българска граматика“ 2021





Научна сесия на Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“ 2022



Състезание по компютърна лингвистика



Състезание по компютърна лингвистика



Шест издания от 2018 г. досега;

Задачите се разработват съвместно от специалисти от Секцията по компютърна лингвистика, Института по математика и информатика и ръководители на отборите в Олимпиадата по лингвистика.







Партньорство



В науката



HUNGARIAN RESEARCH
CENTRE FOR LINGUISTICS



University of
Sheffield



DET DANSKE SPROG-
OG LITTERATURSKAB



ÖAW
AUSTRIAN
ACADEMY OF
SCIENCES



University of
Zurich



UNIVERSIDADE
DE VIGO



University of
Belgrade



UNIVERSITY
OF LATVIA



SAPIENZA
UNIVERSITÀ DI ROMA



Basque Center for Digital Humanities



University of Ljubljana



Wrocław
University of
Science and
Technology



The Árni Magnússon Institute
for Icelandic Studies



Ústav pro jazyk český
Akademie věd České republiky



Sveučilište u
Zagrebu



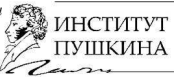
L-Università
ta' Malta



NUI Galway
OÉ Gaillimh



INSTITUUT VOOR
DE NEDERLANDSE
taal/



Deutsches Forschungszentrum
für Künstliche Intelligenz
German Research Center for
Artificial Intelligence



KØBENHAVNS
UNIVERSITET



V.V. Vinogradov
RUSSIAN
LANGUAGE
INSTITUTE
of the Russian
ACADEMY
OF SCIENCES

Språkrädet



NTNU

Norwegian University of
Science and Technology



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



UNIVERSITY
OF ŁÓDŹ



UNIVERSITY
OF ŁÓDŹ



Institute for Computational Linguistics
«Antonio Zampolli»
NATIONAL RESEARCH COUNCIL OF ITALY



Charles
University



UNIVERSITÉ DE
FRANCHE-COMTÉ
600 ans d'histoire



CLARIN
ERIC
Common Language Resources and
Technology Infrastructure



REAL
ACADEMIA
ESPAÑOLA



PRIFYSGOL
BANGOR
UNIVERSITY



UNIVERSITÄT
WIEN

universität
wien



University of Antwerp
CLIPS | Centre for Computational
Linguistics and Psycholinguistics



С бизнеса



CROSSLANG



А DATA PRO



Награди



Награди



Първа награда и грамота от Фонд „Научни изследвания“ към Министерството на образованието и науката за постигнати научни резултати при разработката на научноизследователския проект *Глаголна семантика – проблеми на интерфейса*, 2005 г.





Награди



Цветана Димитрова – Награда за млад учен „Професор Марин Дринов“, 2011 г.

Ивелина Стоянова – Грамота „Джон Атанасов“ за прилагане на научни постижения в практиката и за проекти с висока обществена значимост, 2012 г.

Светлозара Лесева – Награда за млад учен на Института за български език, 2013 г.





Награди



Тодор Лазаров – Награда на БАН за ръководител на най-успешен проект към Програмата за подпомагане на млади учени и докторанти в БАН, 2017 г.

Виктория Петрова – Диплома за второ място от Втория интердисциплинарен докторантски форум, 2019 г.

Валентина Стефанова – Награда за млад учен на Института за български език, 2022 г.



Награди



Светла Коева – Плакет и грамота на БАН за ръководство на проекти на обща стойност над 1 млн. лева през последните пет години, 2014.

Грамота на БАН за значима монография с национално значение за *Semantic Network with a Wide Range of Semantic Relations*, 2020 г.



Награди



2011, 2012, 2013 Най-значим международен проект: Central and South-east Europe Language Resources (CESAR); Applied Technology for Language-Aided CMS (ATLAS)

2015 Най-значимо научноприложно постижение: Българската лексикалносемантична мрежа БулНет
<http://dcl.bas.bg/bulnet/>

2015, 2016 Най-значим международен проект: European Language Resources Coordination (ELRC)

2017, 2018 Най-значим международен проект: CEF Automated Translation for the EU Council Presidency

2020 Най-значимо научноприложно постижение: Кулинарен уърднет
<http://dcl.bas.bg/bulnet/>

2021 Най-значим международен проект: European LEXicographic InfraStructure (ELEXIS)

2023 Монография с национално значение: *Онтология на ситуациите за състояние – лингвистично моделиране.*



Награди



*Онтология на ситуациите за
състояние – лингвистично
моделиране. Съпоставително
изследване за български и руски*

**2022 Най-значим международен проект
на Института за български език**

**2024 Грамота от Фонд „Научни
изследвания“ към Министерството на
образованието и науката за успешна
реализация на проекта**