
SEMANTIC ANALYSIS AND LEXICAL DESCRIPTION OF ARGUMENTS IN PREDICATE-ARGUMENT STRUCTURE IN NATURAL LANGUAGE PROCESSING

Mariya Zhekova

University of Plovdiv “Paisii Hilendarski”, Bulgaria, mimijekowa@abv.bg

Abstract: The study focuses on the ability of a computer to interpret natural human language, to process, retrieve and analyze data obtained from a user's natural language request. The Bulgarian language presents serious challenges to developers and researchers in the field of natural language processing. The purpose of this article is to describe some of these challenges and to present some solutions that would target current and future natural language professionals. The article focuses on the problem of classification and annotation of logical predicates in order to obtain a modern semantic analyzer in natural language processing in Bulgarian. Conceptualization involves the description of many objects and concepts, knowledge about them, hierarchies of types, specifying classes, and the relationships between them. Each sentence, process, or phenomenon in the world is characterized by well-defined relationships between participants in it. In the paper, we present the semantic relations that participants engage in situations described by predicate - argument structures. The development of linguistic analysis aims to determine semantically and syntactically the type of linguistic units and to build upon it by adding an additional level of abstraction over syntactic analysis. The study set up the base of upgrading the existing semantic syntactic resource FrameNet by developing structural accumulative frame models (AFM) for compatibility of knowledge bases, by modeling as accurately as possible to recognize predicate arguments in a predicate-argument frame structure.

Keywords: semantic analysis, natural-language processing, information retrieval, semantic description, semantic analyzer, predicate-argument structure

СЕМНИЧЕН АНАЛИЗ И ЛЕКСИКАЛНО ОПИСАНИЕ НА АРГУМЕНТИТЕ В ПРЕДИКАТНО-АРГУМЕНТНА СТРУКТУРА ПРИ ОБРАБОТКА НА ЕСТЕСТВЕН ЕЗИК

Мария Жекова

Пловдивски университет “Паисий Хилендарски”, България, mimijekowa@abv.bg

Резюме: Изследването се фокусира върху способността на компютъра да интерпретира естествен човешки език, да обработва, извлича и анализира данни, получени от заявка на естествения език на потребителя. Българският език представя сериозни предизвикателства пред разработчиците и изследователите в областта на естествено езиковата обработка. Целта на тази статия е да опише някои от тези предизвикателства и да представи някои решения, които биха насочили настоящите и бъдещите специалисти в областта на естествено езиковата обработка. Фокусираме върху проблема с класификацията и анотацията на логически предикати с цел да получим съвременен семантичен анализатор при естествено езиковата обработка на български език. Концептуализацията предполага описание на множество от обекти и понятия, знания за тях, йерархии на типове, специфициращи класове и релации между тях. Всяко изречение, процес или явление от света се характеризира с точно определени релации между участниците в него. В разработката представяме семантичните релации, с които се свързват участниците в ситуации, описани чрез предикатно - аргументни структури. Лингвистичният анализ в разработката има за цел семантично и синтактично да определи вида на езиковите единици и да надгради като добави допълнително ниво на абстракция над синтактичния анализ. Изследването поставя основите за надграждане на съществуващия семантико-синтактичен ресурс FrameNet чрез разработване на структурни акумулативни фреймови модели (АФМ) за съвместимост на бази от знания, чрез възможно най-точно моделиране с цел разпознаване на аргументите на предиката в предикатно-аргументна структура на фреймите.

Keywords: семантичен анализ, естествено-езикова обработка, извличане на информация, семантично описание, семантичен анализатор, предикатно-аргументна структура

1. ВЪВЕДЕНИЕ

Всяко изречение, процес или явление от света се характеризира с точно определени релации между участниците в него. В разработката ще представим семантичните релации, с които се свързват участниците в ситуации, описани чрез предикатно-аргументни структури. Лингвистичният анализ в разработката има за цел семантично и синтактично да определи вида на езиковите единици и да надгради като добави допълнително ниво на абстракция над синтактичния анализ.

Изследването поставя основите за надграждане на съществуващия семантико-синтактичен ресурс FrameNet чрез разработване на структурни акумулативни фреймови модели (АФМ) за съвместимост на бази от знания, чрез възможно най-точно моделиране с цел разпознаване на аргументите на предиката в предикатно-аргументна структура на фреймите.

Проучването ми има за цел конкретно да определи аргументната структурата на предикатите, съставени от глагол, субект и комплементи. Такъв анализ е необходим от една страна за постигане на автоматичен анализ на нелесните за разпознаване глаголни фрази, а от друга страна, за да може предикатите от този тип в последствие да бъдат съотнесени към бази от данни, описващи АФМ.

Защо точно глагол?

Защото формулировката на всеки процес, явление, състояние започва с глагол, който характеризира действието. Предикатът в изречението е тази част, която променя подлога по един или друг начин. Тъй като подлогът е човек, място или нещо, за което изречението се отнася, предикатът трябва да съдържа глагол, изразяващ действието или състоянието на подлога. В разработката са разгледани съставни предикати, даващи подробности за един и същ субект или явление.

2. ЛЕКСИКАЛНО-СЕМАНТИЧНИ БАЗИ ОТ ЗНАНИЯ

Разгледани бяха компютърни системи като:

- **WordNet** - огромна online лексикално-семантична база от данни, наподобяваща синонимен речник, базирана на семантиката на фреймите и подкрепен с корпус от предварително анотирани голям брой примери. WordNet е форма на традиционна семантична мрежа. Възлите в нея са множества от действителни думи в английския език, които имат общо значение и са синоними в даден контекст. Дъгите изразяват семантични, морфо-семантични и екстралингвистични релации между възлите (BulNet site).
- **BulTreeBank** Wordnet, Българският WordNet (БулНет) - част от многоезична семантична мрежа, съдържа повече от 80 000 синонимни множества, разпределени в девет части на речта - съществителни, глаголи, прилагателни, наречия, местоимения, предлози, съюзи, частици и междуметия. Изградена е от множества от думи, наречени синсети (synsets – synonpmy sets). Всяко синонимно множество - synset, кодира релация на еквивалентност между няколко единици - literals, които имат уникално лексемно значение, принадлежат към една и съща част на речта и изразяват еднакво значение (WordNet site).
- Проектът **FrameNet** – лексикална база данни за английски език. Съдържа над 200 000 ръчно анотирани изречения, свързани с повече от 1200 семантични рамки, предоставя набор от данни за етикетирание на семантични роли, които се използват в приложения като извличане на информация, машинен превод, разпознаване на събития, анализ на настроеността и т.н. Лексикалната база данни представлява семантична мрежа, в която възлите са фреймове или фреймови елементи. Фреймите описват събития, отношения, процеси и явления, заедно с участниците в тях. ФреймНет има за цел описанието на глаголите, като са включени техните задължителни и допустими обкръжения. Указват се също така семантичните и синтактичните ограничения върху обкръженията (Стоянова, 2012а).

3. ФРЕЙМ ЗА ОПИСАНИЕ НА ЗНАНИЯ ЗА СВЕТА

Ще стъпим на теорията на Мински, който описва фреймът като средство за представяне на знания (Мински, 1974), състоящ се от подредени елементи, наречени слотове. Всеки слот съдържа определена информация и носи класификационна информация за синтактичната функция, семантичната роля и принадлежността на езиковата единица, намираща се в него, към част на речта.

Целта ни е да създадем компютърни фреймови модели от анотирани логически форми, в които да поставим ситуации и обекти от действителността – хора, организации, процеси, явления, фрази, описани граматично.

Акумулативният фреймов модел (съвкупността от свързани помежду си фрейми) е едновременно форма, метод и средство за обучение, логически и последователно наредени структури от знания, съставени от обобщени структурни модели – празен фрейм прототип и техните конкретни, попълнени представители - фрейм екземпляр (Zhekova&Totkov, 2019). В така описания модел намират място както описателните, така и базовите знания за всяка предметна област.

Фреймът прототип задава обобщената структура, чрез която се описват екземплярите от даден клас или подклас. Класът е понятие от предметната област (ПО). Фреймът екземпляр представлява фрейм прототип, попълнен с конкретни стойности за всеки конкретен елемент от класа или подкласа (Жекова§Тотков, 2019). Фреймите за представяне на знания имат следните общи белези: релативност, йерархичност, ключови думи, многофункционалност, скелетната форма (наличие на рамка с празни прозорци), асоциативни връзки. Всеки процес във фреймовия модел получава адекватно непротиворечиво представяне на база на участниците в него.

В основата на опорно-фразовата граматика стои схващането, че езикът се състои от езикови обекти (Осенова§Симов, 2007) от вида:

- Езиков израз, който отговаря на отношението (било то признак, отношение, действие, състояние, свойство) - съответства на предиката от тази структура, най-често това са прилагателни или глаголи.
- Езикови изрази, които отговарят на задължителните участници са аргументите и най-често това са съществителни имена или фрази.

Целта на класифицирането на данните във фрейми е да се поддържат и съхраняват данните, да се дефинират връзките помежду им и да се улесни достъпът до тях. По този начин общата обвързаност, йерархичност и подчиненост на действията допринасят за пълното описание на логическия модел на конкретната ПО. Изясняването на основните принципи за класифициране на единиците и създаването на единни и непротиворечиви категории е необходимо условие за приложимостта на класификацията (Стоянова, 2012b). Така структурирана базата от знания се свежда до задача за класифициране, където в отделни фрейми са поместени отговорите на най-често задаваните въпроси за разглежданата ПО.

В теоретичния фреймов модел като елементи на разглежданата фреймова система участват различни понятия (concepts) с техните връзки, описани чрез факти и правила от съответната ПО (Лесева, 2012). В структурно отношение той представлява системно обвързани понятия с различно равнище на абстрактност и е първообраз на бъдещата база данни. Концептуализацията предполага описание на множество от обекти и понятия, знания за тях, йерархии на типове, специфициращи класове и релации между тях. След което ги съпоставяме към аргументите на логическите предикати, в които взимат участие.

Фреймът е модел на съждение, мисъл, изречение, който е съставен от слотове, в които има лексеми или езикови словосъчетания. Вземайки под внимание синтактичните отношения между слотовете, от които е изграден един фрейм и граматическите връзки между тях стигаме до понятието „части на фрейма”. Или части на фрейма са онези слотове, в които има думи или словосъчетания с лексикално-граматичен състав, които влизат помежду си в определени синтактични отношения и се свързват чрез принципа на подчинението.

В изследването е обособено отделно място на частите на речта и техните семантико-синтактични функции, за по-лесно разбиване на фреймовия модел на съставните му части и разглеждането на формалните отношения между тях. В табл.1 е представена структурата на фрейм, неговото съдържание, представяне и организация на езиковите средства в базата му от данни, смисловото му съдържание, логическия му строеж и граматическите връзки между елементите му.

Модел на фрейм – прототип „Понятие“ за формиране на лексикална база знания

Табл. 1. Формален строеж на фрейм-прототип

Име	Тип	Ключ
ID	Идентификатор	PK
Име на фрейм / FrameName	Текстово поле	
Име на фрейм на английски език / FrameNameEN	Текстово поле	
Речниково описание / Description	Текстово поле	
Въпросителна дума / QuestionWord	Текстово поле	
Граматична категория / PartOfSpeech	ID	FK
Семантичен клас / SemClass	ID	FK
Варианти/Словоформи / WordOptions	Текстово поле	
Синонимен ред / Sinonims	Текстово поле	
Корен/ядро на думата / Core	Текстово поле	FK
Представка / Prefix	Текстово поле	FK
Наставка / Suffix	Текстово поле	FK
Семантични ограничения / SemLimitations	ID	FK
Наследява се от / SubFrameOf	ID	FK
Е аргумент във / IsArgumentIn	ID	FK

4. СЕМАНТИЧЕН АНАЛИЗ НА НИВО ФРЕЙМ

Автоматичната обработка на информацията изисква машинно разбираемо представяне на нейната семантика. Чрез семантичното обозначаване на лексикалните единици се извършва по-задълбочен анализ на постъпилния текст, който се базира на синтактичен анализ, надграден с аотиране на семантична функция. Освен семантични анализатора прави и синтактични заключения, защото част от значението на думата е и ролята ѝ в изречението.

Моделирайки мрежови логически структури от знания, които ще бъдат източниците, от които по-късно ще извличаме информация, ние всъщност структурираме семантичните си възприятия чрез смесване на понятията около т.нар. семантични прототипи. Те се състоят от пространствено групирани възли с ключови думи, представляващи понятия и свързващи линии-дъги, представляващи семантичната връзка между понятията.

Във формалния синтаксис естественият език се представя като теоретико-множествена структура от думи (словоформи). Кои множества принадлежат към езика, се определя от математически механизъм, който свързва едно езиково множество с неговата структура (“Автоматична граматична корекция”, 2010).

Обработката на естествен език е област от изкуствения интелект, която изисква компютърните системи да обработват естествен език (Sowah, 2018). Обработката на сложни въпроси включва идентифицирането на няколко вида семантични структури от фрейми. Първо, трябва да разпознаем типа на очаквания отговор, който е богата семантична структура в случай на сложен въпрос, или просто понятие в случай на фактически въпрос. Второ, трябва да идентифицираме класа на въпроса или модела на въпроса. Трето, в случай на сложен въпрос, който е част от работен процес, трябва да открием точно вида на процеса, т.е. да интерпретираме темата на въпроса.

Автоматичната система за отговор на въпрос на естествен език се основава на предварително определен модел на семантична структура на аргументите на предикати, включени в лексикалния речников ресурс или извлечени от корпуси като BulTreeBank, PropBank или FrameNet. Те се обучават с помощта на контролирани методи за машинно обучение като използват знания, означени като предикати и семантични роли на аргументите им (Hartmann, 2017).

В разработката предмет на описание ще бъдат глаголите заради тяхната основна и водеща роля в изречението. А именно:

- глаголът определя предикативно отношение между обектите, описвани от групите в изречението;
- глаголът и неговата семантика определят състава и структурата на изречението.

Следователно основна задача ще бъде намирането на глагола в изречението.

За да бъде анализът по-задълбочен и точен, е необходимо на първо място да се уточнят глаголите със специфичната им аргументна структура, които са обект на изследване, и техните значения, както и възможността им за участие в предикативни конструкции.

Табл. 2. Списък с част от глаголите в разглежданата предметна област – Университет

администрира	внедрява	вода	допълва	закрива
активира	влива	говоря	анкетира	анулира
актуализира	влиза	дава	доразвива	замества
алармира	включва	датира	доставя	заменя
анализира	взема	движи	достига	действа
архивира	взима	декларира	запазва	записва
блокира	върчва	делегира	живее	започва
броя	върща	допринася	забранява	

5. ПРЕДИКАТНО-АРГУМЕНТНА СТРУКТУРА НА НИВО ФРЕЙМ

Формализирайки семантично предикатите ние се стремим да сведем лингвистичните характеристики към математически модели, качествените характеристики да сведем до количествени. Математическата логика е дял от математиката, който дава техники за дефиниране на синтаксиса и семантиката на различни формални езици и изследване на свойствата на тези езици. Предикатната логика разкрива логическата форма на твърденията в детайли и за нея е известно, че тя е дала математическата основа на релационните бази данни. В нейната същина влизат константи и променливи, с които означаваме имена на обекти и предмети от разглежданата предметна област, техните свойства и отношения.

Предикатно-аргументната структура е пренесена от предикатната логика, където отношенията между участниците в ситуацията от действителността се изразяват чрез пропозиционални функции (P(x), P(x,y), P(x,y,...)) (Несторова, 2008). С P се отбелязва предиката.

Ще представим базата данни като съвкупност от предикатни структури, дефинирани изчерпателно с термините на компютърната лингвистика.

Предикацията е приписване на признак на предмет. Различните признаци (свойства или отношения) се наричат предикати, а предметите, към които те се отнасят, са аргументи на предиката. Разглеждани като части на речта, аргументите влизат в семантични отношения с глагола. Едно и също съществително може да играе различна семантична роля в зависимост от семантиката на глагола, на който е аргумент, а не от собствената си семантика. Чрез значението на съществителното се решава дали то може да бъде аргумент на дадения глагол, но чрез значението на глагола се решава каква роля може да има то именно като аргумент, а не като дума.

Например:

Идентифицира се (*име, ЕГН*)

Студент е (*име, фак. номер, специалност, университет, курс, форма на обучение*)

Живее (*име, град, адрес, сграда, етаж, от, до,*)

Роден е (*име, дата, град*)

Примери за клас-подклас:

Табл.3. Наследяване на характеристиките, дефинирани в слоговете на фрейм-прототипа

Прототип	Екземпляр
Студент	Студент 1
<i>A kind of (от клас Младеж)</i>	<i>A kind of (от клас Младеж)</i>
<i>Университет(Име)</i>	<i>Университет(ПУ Паисий Хилендарски)</i>
<i>Специалност()</i>	<i>Специалност(Инженерна физика)</i>
<i>Име(Име, Фамилия)</i>	<i>Име(Кристиан Кръстев)</i>
<i>Пол()</i>	<i>Пол(мъжки)</i>
<i>Възраст()</i>	<i>Възраст(27)</i>

В структурата на изречението $S = VP + NP$, VP е група на глагола, а NP е група на съществителното. Аргументите, които глаголет изисква като задължителна част от глаголната фраза (VP), която проектира, се наричат “вътрешни”. Аргументът, който се реализира като подлог, се нарича “външен”.

Глаголет като лексикален клас се субкатегоризира на две големи групи според способността на глаголите да бъдат предикати или не. Глаголите-връзка (*link verbs*) свързват подложната фраза с именни фрази вътре в предиката, които носят семантиката му. Пълнозначни (*full verbs*) са глаголите, които могат да бъдат предикати. Семантиката на един глагол определя броя на участниците в описваната ситуация, броя на семантичните роли и, отгук, броя на глаголните аргументи. Всеки глагол има поне един аргумент. Според броя на аргументите си, глаголите могат да бъдат непреходни (*Intransitive*) – само с един аргумент, реализиран като подлог и без компленти-допълнения – или преходни (*Transitive*) – с повече от един аргумент и с компленти-допълнения (Стамболиева, 2015).

Глаголите играят централна роля за структурата и значението на изреченията. Те образуват естествени класове на базата на различни (множества от) свойства. От важно значение за определяне на аргументната структура на даден предикат е семантичният тип, към който той принадлежи. Според приетата класификация те са четири типа: статични / състояния (*states*); каузативни / процеси (*accomplishments*); инхоативни / събития (*achievements*); активни / действия (*activities*). В WordNet тези категории са залегнали в организацията на семантичните класове, на които са разделени глаголните значения. Глаголите са разделени на подкласове в зависимост от това, дали са за възприятие, за движение, за допир, за изразяване на чувства, за мисловна дейност и т.н., и според синтактичната си рамка.

Потенциалът на глагола да избира своите компленти се нарича валентност на глагола (Schulte, 2009). Комбинацията от функционални компленти, които образуват предикатно-аргументната структура на глагола, ще наричаме глаголна рамка или глаголен фрейм.

В описанието на глаголния фрейм място ще намерят слогове за съставлящите го елементи, кои от тях са задължителни или незадължителни, описание на синтактичните и семантичните им функции. Степента на приемливост по отношение на семантичната реализация на аргументите се променя, следователно глаголет дефинира „селективни предпочитания“ за неговите компленти. Селективните предпочитания се отнасят не само до синтактичната функция на комплемента в рамките на глаголната рамка, но се взема предвид и семантичната роля на съответния аргумент.

Потенциалът на глаголите да избират своите допълнения се нарича подкатегоризация на глаголите или валентност на глагола, а комбинация от функционални допълнения, които се предизвикват от глагол, се нарича глаголна рамка.

6. СЕМАНТИЧНИ РОЛИ НА НИВО ФРЕЙМ

Предикатно-аргументната структура се формализира с помощта на фрейм и елементите, участващи в състава му. Отделните аргументи и отношенията между тях се описват лингвистично и чрез семантични роли. Целта на семантичните роли е да обобщят сходни семантични свойства на класове от аргументи.

Филмор е първият автор на идеята за семантичните падежи 1968-1971г. Той предлага седем роли – агентив, инструменталис, датив, фактитив, локатив, обектив, бенефактив, но този списък се допълва от неговите последователи. Според Джакендорф семантичните роли са част от концептуалната структура, а не част от синтаксиса. Граматичните категории могат да имат множество различни семантични роли (Несторова, 2006). Предикатно-аргументната структура (ПАС) е представянето на един фрейм и елементите, участващи в състава му, описана от предиката чрез аргументите му и отношенията, описани чрез семантични роли. Целта на семантичните роли е да обобщят сходни семантични свойства на класове от аргументи. Семантични (тематични, тета) роли са разгледани от Й. Пенчев (1999г).

- AG – агенс – реален или потенциален деятел с признаци, одуш.лице
- AC – актор – реален или потенц.вършител на действие с признаци, одуш.
- ELEM – елементатив – източник на действието с признаци, неодуш
- DOER – реален или потенц.изпълнител на действието с признаци, неодуш, артефакт
- SUBE – субект на физ.състояние
- EXP – експериментатор – реален или потенц.субект на физиол.или псих.състояние или мент.дейност
- SUBC – субект на сравнение
- OVC – обект на сравнение
- OB – обект – реален или потенц.обект, върху който се разпространява действието
- PAC – пациенс – реален или потенц.обект, върху който се разпространява действието
- ADR – адресат – адресатът на съобщение с признак, одуш.
- PERC – перцептив – субект на възприятие или физ.отражение при глаголи
- INST – инструмент – оръдие за действие с признак, неодуш
- MED – медиатив – средство за действие с признак, неодуш
- RES – резултатив – резултат от извършеното действие
- LOC – локатив – място, на което се извършва действието

Модел на фрейм – прототип за представяне на аргументна структура на предикати

Глагол	Аргумент 1	Аргумент 2	Аргумент 3	Аргумент 4	Аргумент 5	Аргумент 6
	<i>Агент</i>	<i>Експериментатор</i>	<i>Тема</i>	<i>Резултатив</i>	<i>Темпоратив</i>	<i>Локатив</i>
<i>ИЗПИТВА</i> <i>ОЦЕНЯВА</i>	Изпитващ Преподавател Кой?	Изпитван Студент Кой?	Дисциплина Изучаван предмет Какво?	Оценка Резултат Колко?	Дата на изпита Кога?	Място аудитория, комп.зала, сем.зала Къде?
<i>ПОДАВА</i> <i>ДОКУМЕНТ</i>	Подател Кой?	Вид документ Какъв?	Номер документ Кой?	Служител приел документа Кой?	Отдел Кой?	Статус на документ Какъв?

Голяма част от ПАС са именно глаголи, които са типични предикати и това обяснява тяхното централно място в изречението. Глаголите се различават структурно по броя и типа на аргументите, които могат да приемат.. Те ограничават семантичните свойства на аргументите си (селекционни ограничения). "Еквивалентните" аргументи имат еднакви семантични свойства в различни синтактични реализации и различни предикати. Според когнитивния подход семантичното представяне на даден глагол е равносилно на представяне на концепцията за събитието, което глагола кодира (Славчева, 2011).

7. ЗАКЛЮЧЕНИЕ

Предикатно-аргументните структури отговарят на ситуации от реалната действителност, следователно тяхното описание е еквивалентно за всички езици. Естествено-езиковия интерфейс за информационна система от тип въпрос-отговор е разработен като начало на български език, като въпроса за разработването на други езици е оформен като задача за следващи научни разработки.

Остава да бъдат проучени много въпроси. Например: Колко голяма е извадката от понятия, необходима за описване знанията за една предметна област? Колко стабилна е структурата от концептуални модели? Колко заменяеми са техниките за създаване на фрейм-прототипи, които се използват в тестовата система? Изследването продължава и с анализа на ефективността на отговорите, които системата връща на зададените потребителски въпроси.

ЛИТЕРАТУРА

- Жекова, М. & Тотков, Г. (2019). *Фреймов модел за представяне на семантични роли и процеси за изграждане на естествено-езиков интерфейс*. Сборник от XII Национална конференция „Образованието и изследванията в информационното общество“. Пловдив
- Лесева, Св. (2012). *Лексикална семантика и синтактична реализация на аргументите*. Секция по компютърна лингвистика БАН. София
- Колектив от секцията по Компютърна лингвистика от Института за български език при БАН. (2010). *Автоматична граматична корекция на българския език*
- Несторова, П. (2006). *Семантични роли на подлога в съвременния български език*. Пловдив
- Несторова, П. (2008). *Семантични и синтактични особености на предикати в българския език*. Пловдив
- Осенова, П. & Симов, К. (2007). *Формална граматика на българския език*. София
- Славчева, М. (2011). *Семантични дескриптори на рефлексивни по форма глаголни структури в съвременния български, френски и унгарски език*. Институт по паралелна обработка на информацията БАН
- Стамболиева, М. (2015). *Класове думи: Английски-български паралели*. Годишник на Департамент Англицистика. НБУ
- Стоянова, Ив. (2012). *Автоматично разпознаване и тагиране на съставни лексикални единици в българския език*. София
- Hartmann, S. (2017). *Knowledge-based Supervision for Domain-adaptive Semantic Role Labeling*. Darmstadt
- Minsky, M. (1974). *A Framework for Representing Knowledge*. MIT. Cambridge
- Schulte, S. (2009). *The induction of verb frames and verb classes from corpora*. Corpus Linguistics. Berlin
- Sowah, E. (2018). *Natural language processing in cooperative query answering databases (NLPiCQA)*. Computer Science and Technology. China
- Zheкова, М. & Тотков, Г. (2019). *Conceptual Frame Model For The Presentation Of The Concepts And Rules In Natural Language Interface For Database*. IOP Conference Series: Materials Science and Engineering. Plovdiv
- BulNet. <https://dcl.bas.bg/bulnet/>
- WordNet. <https://wordnet.princeton.edu/>