

Лексичка база Фрејмнет: неколико примера оквира из домена ризика

УДК 81'322.2

САЖЕТАК: У раду се даје кратак приказ теорије семантике оквира (енгл. *Frame Semantics*), на којој је заснована лексичка база Фрејмнет (енгл. *FrameNet*). Представљена је концепција ове мреже, као и могућности њене примене. Представљена је и лексичка анализа која се примењује у пројекту израде Фрејмнета и указано на разлике између анализе засноване на оквиру у односу на анализу засновану на речи. Затим је приказано неколико повезаних оквира које призивају речи из домена ризика. У раду је представљена и платформа NLTK (енгл. *Natural Language Toolkit*), помоћу које се могу користити разни језички ресурси, међу њима и Фрејмнет. Завршно поглавље пружа анализу именице *ризик* на корпусу рударства. Представљени су најчешћи колокати ове именице, скица њене употребе, конкорданце за поједине моделе, проналажење синонима и повезаних речи у виду тезауруса, графички приказ фреквенција појединих колокација, као и облака речи.

КЉУЧНЕ РЕЧИ: Српски језик, семантика оквира, Фрејмнет, сценарио ризика, корпус рударства, обрада природних језика.

РАД ПРИМЉЕН: 15. јули 2021.

РАД ПРИХВАЋЕН: 6. септембар 2021.

Александра Марковић

aleksan-

dra.markovic@isj.sanu.ac.rs

Институт за српски језик

САНУ

Београд, Србија

Ранка Станковић

ranka.stankovic@rgf.bg.ac.rs

Универзитет у Београду

Рударско-геолошки факултет

Београд, Србија

Наталија Томић

ntomic@hotmail.com

Универзитет у Београду

Београд, Србија

Оливера Китановић

olivera.kitanovic@rgf.bg.ac.rs

Универзитет у Београду

Рударско-геолошки факултет

Београд, Србија

1. Увод

Теорија семантике оквира (енгл. *Frame Semantics*) је когнитивносемантички приступ Чарлса Филмора који експлицитно

повезује значења неке речи са синтаксичким контекстима у којима се та реч јавља (Atkins, Fillmore, and Johnson 2003, 254). Анализом значења речи баве се углавном лексикографи, као и они који се баве семантиком. Међутим, уколико је циљ представљање начина на који се нека реч заиста користи, анализа података из корпуса показује се као прилично компликован задатак, с обзиром на број конкорданци које савремени корпуси нуде за поједине кључне речи. Теорија семантике оквира, према оцени више аутора (Atkins 1994; Gildea and Jurafsky 2002; Atkins, Fillmore, and Johnson 2003; Pradhan et al. 2005; Boas and Dux 2017; Jurafsky and Martin 2020), представља поуздан, научно утемељен начин да се анализира и опише начин употребе неке речи. У основи овог приступа лежи идеја да се свако искуство које памтимо јавља у неком смисленом контексту и да га можемо и упамтити баш захваљујући томе што имамо когнитивну схему или оквир за разумевање искустава. Филмор сматра да се речи уче у таквим смисленим контекстима, а ти су контексти неопходни и за процес разумевања, када у сећању призивамо искуства кроз која смо научили поједине речи. Оквир идентификује тип искуства, даје му структуру и целовитост, односно даје значење предметима, догађајима и односима у том искуству (Fillmore 1976, 26).¹

1.1 Концепција мреже Фрејмнет

Фрејмнет² је лексичка база података енглеског језика, заснована на анотацији примера употребе неке лексичке јединице (у даљем тексту ЛЈ; енгл. *lexical unit*) у аутентичним текстовима (дакле, у корпусу, а не у конструисаним примерима). Основна идеја своди се на то да се значења већине речи најбоље могу разумети на основу семантичког оквира, појмовне структуре налик на сценарио, који представља опис типа ситуације, догађаја, односа или ентитета, као и учеснике тих ситуација, догађаја и односа (Ruppenhofer et al. 2016, 7). На пример, ризиковање у типичном случају подразумева следећи скуп појмова: особу која је централна за сценарио РИЗИКА – *Протагонисту*, који свесно или несвесно доноси одлуку да ризикује или подлеже неком

1. Термин *оквир* Филмор користи као општи термин за скуп појмова који су у литератури о разумевању природних језика познати под називима *схема*, *сценарио*, *когнитивни модел*, *народна теорија* и неки други (Fillmore 1982, 111).

2. Пројекат се развија у Међународном институту за рачунарске науке у Берклију (International Computer Science Institute, Berkeley) од 1997. године.

ризику; могући *Лош исход* или штету; *Одлуку* која може довести до лошег исхода; *Циљ* који се жели постићи; *Околности* у којима се ризик јавља и *Вредност*, особу или предмет који су драгоцени Протагонисти и који су угрожени у датој ризичној ситуацији (Fillmore and S. Atkins 1994, 367).

1.2 Лексичка анализа заснована на оквиру

Лексичка анализа заснована на теорији семантике оквира подразумева анализу семантичког садржаја неке лексичке јединице, идентификацију њеног семантичког суседства, откривање граматичких конструкција у којима се јавља у корпусу, као и бележење свих конструкција у којима дата ЛЈ изражава пун семантички потенцијал. Сматра се да је неопходно описати све допуне и одредбе у тим конструкцијама. Посебна пажња поклања се речима које се не могу правилно употребити уколико се не знају конструкције у којима се јављају. Такве речи називају се речима које призивају оквире³ (енгл. *frame-evoking words*), а то су на првом месту глаголи, али и именице, придеви и прилози (Atkins, Fillmore, and Johnson 2003, 252).⁴

Основне јединице лексичке анализе у Фрејмнету су оквир и ЛЈ, под којом се подразумева лексема у једном од својих значења (Fillmore et al. 2003, 297), (Ruppenhofer et al. 2016, 7).⁵ За разлику од онога што је уобичајено у лексикографији, фокусирања на једну реч, односно лексеми, и истраживања свих њених значења, тј. ЛЈ, у Фрејмнету се

3. Исти еквивалент енгл. глагола *to evoke* користи се и у преводу рада Fillmore 1982 у (Rasulić and Klikovac 2014, 79).

4. Теорија семантике оквира подстакла нас је да укажемо на потребу да се за четири основне класе речи које призивају оквире (именице, придеве, глаголе и прилоге) наводе релевантне конструкције у описним реченицама српског језика (Марковић 2017, 34–41).

5. У српској лексиколошкој литератури, као и у синтаксичким радовима који се баве односом граматике и речника, користи се различита терминологија за оно што се у Фрејмнету назива лексичком јединицом (нпр. у универзитетском уџбенику лексикологије под лексичком јединицом подразумева се одредница или лема (Драгићевић 2007, 30)), док Поповић (2003, 202–203) указује на важност тога да се пажња у синтакси поклања конкретним лексемама, односно лексемама употребљеним у једном од више значења, које назива сублексемама. Ми смо се у овом раду определили за термин лексичка јединица, у складу с терминологијом мреже и приступа који представљамо.

у истом тренутку анализирају ЛЈ које припадају истом оквиру⁶ (Fillmore et al. 2003). Тако су, на пример, у односу на оквир *Бити угрожен* (енгл. *Being at risk*) дефинисане именице *ризик*, *опасност*, *безбедност*, *рањивост*, придеви *несигуран*, *безбедан*, *сигуран*, *поуздан*, *рањив* и др.

Процес описивања ЛЈ у Фрејмнету представљен је у (Fillmore et al. 2003).⁷ Он започиње неформалним описом оквира ком та јединица припада, прецизније описом типа ситуације или збивања представљених оквиром и стварања листе речи које се могу објаснити повезивањем с датим оквиром (299).⁸ Затим се бира циљна ЛЈ, централни члан оквира (у терминологији Фрејмнета *target*), ЛЈ у односу на коју се ради анотација; то је типично једна реч, али може бити и вишечлана реч или фразеологизам (Ruppenhofer et al. 2016, 21) и испитује њена употреба екстраховањем из корпуса реченица које је садрже.

Свој увид у значења изабране лексеме, добијен на основу посматрања примера из корпуса, лексикограф који ради на мрежи Фрејмнет пореди са значењима дате лексеме у неком од референтних речника.⁹ Када стекне јаснију представу о значењима, покушава прецизније да опише оквир у који спада изабрана ЛЈ. Затим пише дефиницију оквира – схематски опис типа ситуације која лежи у основи неке речи, заједно с називима улога учесника описане ситуације, које се називају елементима оквира (енгл. *frame elements*). У лексикографском смислу важан је начин на који елементи оквира добијају језички израз у реченицама у којима се јавља посматрана ЛЈ (Fillmore et al. 2003, 304–305).

6. Овде је у питању разлика између два приступа издвајању значења, први је заснован на речи (*word-based*), а други на оквиру (*frame-based*) (Atkins, Fillmore, and Johnson 2003, 254).

7. Иако се процес описује као низ одвојених корака, који се одвијају одређеним редом, аутори напомињу да се заправо у сваком тренутку може вратити на неки од претходних корака и ревидирати нека од одлука (Fillmore et al. 2003, 299).

8. Тај опис подразумева: 1) схематски опис типа ентитета или ситуације представљене оквиром; 2) бирање описних ознака за обележавање елемената оквира; 3) писање радне листе речи које припадају датом оквиру (припадност лексичких јединица истом оквиру указује на то да реченице које садрже те ЛЈ подлежу сличној семантичкој анализи) (297).

9. Пошто су анализи подвргли дефиниције глагола *ризиковати* у десет општих речника енглеског језика, Филмор и Еткинсова закључили су да чак ни речници сличних величина и намена не издвајају основна значења овог глагола, која припадају основном речничком фонду (Fillmore and S. Atkins 1994, 353).

1.3 Елементи оквира

Елементи оквира често се могу посматрати као примери општијих семантичких улога (као агенс, доживљивач, пацијенс), али се дефинишу на начин који је специфичан за поједине оквири (енгл. *frame-specific*). Постоји више разлога за то, а најважнији је тај што омогућава да се направе прецизне дефиниције елемената оквира за одређену групу речи, без потребе за утврђивањем начина на који ће се оне уклопити с малим, унапред задатим скупом општијих семантичких улога (305).

Најпре се утврђују централни елементи оквира (енгл. *core frame elements*).¹⁰ Централни елементи су они који представљају обавезне компоненте оквира и који неки оквир чине јединственим и различитим од других. Поред централних, постоје и периферни елементи, који се могу појавити у свим оквирима у којима неки агенс врши неку радњу (периферни елементи означавају појмове МЕСТА, ВРЕМЕНА, НАЧИНА, СРЕДСТВА).¹¹ Дешава се да неки од централних елемената не буде и језички изражен, иако је обавезан у појмовној структури оквира; такав случај назива се нултим појављивањем (енгл. *null instantiation*) и аотира се у бази (320). Пошто се одреде централни и периферни елементи, дефинише се оквир.¹²

Филмор и Еткинсова закључили су, након поменутог анализе глагола *ризиковати* у референтним реченицама енглеског, да се допунама не

10. Постоје неке формалне особине које помажу у одређивању статуса елемента као централног (нпр. централни су они елементи који морају бити изражени, као и они који, уколико остану неизражени, добијају одређену интерпретацију (нпр. у реченици *John arrived* није изражено место на које је Џон стигао, као циљна локација, GOAL, али се то место разуме на основу контекста) (Ruppenhofer et al. 2016, 23–24).

11. Дистинкција централни/периферни елементи оквира одговара у грубим цртама подели на аргументе и одредбе у традиционалној граматичкој анализи (Fillmore et al. 2003, 310). Периферни елементи не могу бити у функцији субјекта или објекта циљног глагола и често су изражени прилозима или предлошко-падежним конструкцијама (319).

12. Ruppenhofer et al. (2016, 65) издвајају и додатне типове елемената оквира: они који се јављају у зависним клаузама имају периферни или екстра-тематски статус (енгл. *extra-thematic*, нпр. ВРЕМЕ, РАЗЛОГ). Осим тога, постоје и елементи којима се приписује ознака *Core-Unexpressed*; то су они елементи неког оквира који се понашају као централни неизражени, али који се можда неће наћи међу елементима оквира потомака датог оквира (24–25). У овом раду нећемо се бавити ни једним ни другим.

поклања довољно пажње (иако су веома битне, посебно у речницима за учење језика), а да постоје и неки други реченични чланови које речници занемарују, а који морају бити јасно издвојени и описани како би се посматрани глагол употребио правилно. На пример, радња коју изводи особа која нешто ризикује (и која може имати различите синтаксичке реализације): Ризиковала је свој живот *покушавајући да спасе дете које се давило*; затим, циљ који онај који ризикује има на уму кад се излаже ризику: Ризиковала је свој живот *како би спасла мој* (Fillmore and S. Atkins 1994, 362). Радња којом неко нешто ризикује спада у централне елементе оквира, док циљ ради ког се ризик предузима спада у периферне.

1.4 Односи између оквира – мрежа оквира

Пошто се дефинишу оквир и његови елементи, дати оквир се повезује са другим оквирима. На тај начин оквири, њихови елементи и с њима повезане ЛЈ добијају место у семантичком простору (Ruppenhofer et al. 2016, 79) и чине мрежу. Успостављање односа између оквира омогућава уочавање и бележење семантичких генерализација на основу типова учесника, догађаја и сл. Оквир може бити повезан са оквирима које разрађује или наслеђује, са онима који су његови подоквири, као и са онима које користи. Односи између оквира су усмерени или асиметрични: наиме, апстрактнији и независнији оквир се назива *Над_оквиром* (енгл. *Super_frame*) оквира који је зависнији и мање апстрактан и који се назива *Под_оквиром* (енгл. *Sub_frame*) (79).

Дефинисано је неколико врста односа, од којих су најважнији следећи (79–84):

- У односу *Наслеђивања* (енгл. *Inheritance relationship*) подоквир (подређени оквир) представља прецизнију разраду надоквира (надређеног оквира). Сви елементи оквира и подоквири надређеног оквира имају своје парњаке у подређеном оквиру, а он може имати додатне подоквири, елементе оквира и ограничења семантичког типа која се не јављају у Родитељу (Fillmore et al. 2003, 311). Нпр. оквир *Изложити се ризику* (*Run_risk*) наслеђује оквир *Вероватноћа* (*Likelihood*).
- Однос *Коришћења* (енгл. *Using*) постоји када се неки оквир на веома уопштен начин позива на структуру неког апстрактнијег, схематизованог оквира. На пример, оквир *Клађење* (*Wagering*) користи оквир *Изложити се ризику* (*Run_risk*); *Брзина* (*Speed*)

користи оквир *Покрет* (*Motion*); оквир *Брбљивост* (*Volubility*) користи оквир *Комуникација* (*Communication*) (Ruppenhofer et al. 2016, 83).

- *Угао_гледања* (енгл. *Perspective_on*) је релација слична општијој релацији *Користићења*, али повезане оквири ограничава у већој мери (82). Примена ове релације указује на постојање бар два могућа угла гледања на један неутралан оквир. На пример, оквир *Сценарио_ризика* (*Risk_scenario*) је неутралан, док су оквири *Ризична_ситуација* (*Risky_situation*), *Бити_угрожен* (*Being_at_risk*) и *Изложити се_ризички* (*Run_risk*) перспективизовани; оквири *Запошљавање* (*Hiring*) и *Добити_посао* (*Get_a_job*) перспективизују оквир *Почетак_радног_односа* (*Employment_start*) са становишта послодавца, односно запосленог.

Пошто се у базу унесу дефиниције оквира и његових елемената, оквиру на ком се ради додаје се ЛЈ (на пример, оквиру *Бити_угрожен* додаје се ЛЈ *ризик*). Затим се за дату ЛЈ уносе информације о врсти речи, значењу, формалном саставу (једна реч или вишечлани израз). Након тога прецизирају се претраге које ће омогућити да се из корпуса¹³ екстрахују оне реченице (поткорпус) које садрже лексему коју испитујемо (у овом примеру именицу *ризик*) и чија је граматичка форма таква да указује на оно значење дате именице (ЛЈ) које је везано за оквир *Бити_угрожен* (*Being_at_risk*). Циљ је да се из поткорпуса избаце све оне реченице у којима задата кључна реч не представља ЛЈ која се везује за оквир који се креира. Пошто се спецификују претраге важне за дату ЛЈ, низ аутоматских процеса генерише поткорпус спреман за анотацију. Након што се одбаце предуге и друге неадекватне реченице, бира се од три до пет примера за сваки модел – циљ је илустровати разноликост модела, а не постићи статистичку репрезентативност.

Кад се заврши с анотацијом, примењују се алати за испитивање анотираних реченица и валентних модела који су у њима реализовани. Постоје два типа извештаја у виду динамичких веб-страница (*LexUnit Report*, *Lexical Entry Report*; извештаји се аутоматски генеришу пошто се заврши анотација и доступни су и на јавној веб-страни Фрејмнета). Први показује све анотиране реченице за једну ЛЈ и у њему се наводе сви елементи оквира пронађени у актуелном оквиру (табела елемената

13. Fillmore et al. (2003, 304) користе **Британски национални корпус** (енгл. British National Corpus).

оквира); сваки елемент се боји одређеном бојом, а истом том бојом елементи су обележени и у анотираним реченицама. Други представља резиме синтаксичких реализација елемената оквира и валентних модела ЛЈ у две табеле (Fillmore et al. 2003, 326–328).

Пошто се у Фрејмнету бележе и елементи оквира (за оквир специфичне семантичке улоге), као и њихове језичке реализације, за овај опис важни су и појмови *валентне групе*, *валентног модела* и *описа валенце*.¹⁴ Елемент оквира заједно са својом граматичком реализацијом (тип јединице и њена функција у реченици) чини валентну групу; скуп валентних група реализованих у једној реченици чини валентни модел; скуп свих валентних модела које реализује једна лексичка јединица чини опис валенце (Atkins, Fillmore, and Johnson 2003, 255–257).

1.5 Примене мреже Фрејмнет

Фрејмнет је доступан на адреси <https://framenet.icsi.berkeley.edu/>. Може се претраживати и прелиставати онлајн, али и преузети и користити. Као што је на [сајту](#) истакнуто, може се користити у различите сврхе: као речник за учење језика (с обзиром на то да садржи више од 13.000 ЛЈ); као речник валенце; као скуп података за обуку програма који обележавају семантичке улоге,¹⁵ што га чини значајним дигиталним језичким ресурсом за истраживања у области обраде природних језика (садржи више од 200.000 ручно обележених реченица повезаних са више од 1.200 семантичких оквира).

Фрејмнет је настао као лексичка база података за енглески, а касније су развијане базе за друге језике (француски, кинески, португалски, немачки, шпански, јапански и др.), кроз различите и независне пројекте, али уз исти формализам и концепцију развоја. Покренут је [пројекат](#) поравнавања датих база података за разне језике.

14. Својство глагола да за себе „везује“ зависне елементе (аргументе) назива се валенца. У зависности од броја аргумената које везују за себе, глаголи могу бити једновалентни (кад се могу употребити само са субјектом), двовалентни (са субјектом и објектом) итд.

15. У подтачки 1.6 представимо нека од истраживања у којима се на разне начине користе Фрејмнет и програми за бележење семантичких улога у хрватском, словеначком и српском.

1.6 Претходна истраживања

У овом одељку осврнућемо се на истраживања која су се бавила бележењем семантичких улога (анотација семантичким улогама) у српском и њему сродним језицима и испитивањем значења именице *ризик* и глагола *ризиковати*.

У раду Gantar et al. (2018) представљен је модел за обележавање семантичких улога у словеначком и хрватском, развијан у оквиру међународног билатералног пројекта *Semantic Role Labeling in Slovene and Croatian*. Циљ је био развијање ручно анотираних корпуса који би се користио као база података за обуку за системе надгледаног машинског учења. Описан је и експеримент аутоматског обележавања семантичких улога заснован на надгледаном машинском учењу. За сваки корпус представљени су најфреквентнији глаголи, семантичке улоге и типични семантичко-синтаксички обрасци за најфреквентније глаголе. У оба корпуса најфреквентнији је глагол *бити* и семантичка улога пацијенса, а затим агенса (95–96). У раду су семантичке улоге бележене на стабилним семантичко-синтаксичким моделима (96–97), али остаје питање оправданости таквог поступка будући да се семантичке улоге и оквири везују за ЛЈ, дакле за (глаголске) лексеме у једном од значења.

У раду Брач and Анић (2019) описан је пројекат чији је циљ развој методологије за бележење семантичких улога у језику специјализованих домена (конкретно авијације) који би се могао користити за све посебне области. Ауторке разматрају да ли је боље користити општије семантичке улоге или улоге специфичне за поједине глаголе и оквири, што је карактеристика Фрејмнета. Њихов закључак је да превелики број специфичних семантичких улога успорава анотацију а не доприноси значајно бољитку терминолошких ресурса, док се скуп општијих семантичких улога мора незнатно допунити (545).

У раду Wasserscheidt and Hršić (2020) спроведено је интересантно истраживање о томе да ли лексеме које се користе и као термини (конкретно правни термини) и као речи општег лексичког фонда имају различита значења (или призивају различите оквири) у српском и хрватском језику (као језичким варијантама). Идеја је потекла од контрадикције у литератури посвећеној семантици оквира; наиме, у изворној Филморовој варијанти теорије указује се на разлике у оквирима које постоје између појединих говорника, друштвених група и култура; на те се разлике у радовима других аутора заборавља, па се оквири посматрају као универзалне структуре, независне од језика (88–89). Аутори су испитивали да ли *одредба* као правни термин

и реч општег лексичког фонда призива различите оквири у српском и хрватском, а испитивање је спроведено на корпусима ових језика. Коришћена је дистрибуциона анализа, чија је основна идеја да се значење речи може разумети на основу њеног контекста, али је тој анализи подвргнут и сам контекст. У анализи контекста примењена је и теорија семантике оквира (Wasserscheidt and Hrštic 2020, 90). На основу ове двоструке дистрибуционе анализе аутори су закључили да нема значајне разлике у значењу *одредбе* у посматраним корпусима, као и то да се метод двоструког кластерованја може користити у сложеној семантичкој анализи, која се пак може уклопити у појмовне структуре Фрејмнета¹⁶ (108).

Иако није од непосредног значаја за наш рад (јер представљена анализа није заснована на теорији семантике оквира), поменућемо истраживање у ком се најпре примећује да је ризик постао важна тема истраживања у друштвеним наукама, али да се, када се говори о ризику, о значењу саме речи *ризик* не зна довољно (Hamilton, Adolphs, and Nerlich 2007, 164). Аутори истраживања, вођени поменутиим запажањем, подвргавају анализи значење именице *ризик* и глагола *ризиковати* у корпусу разговорног језика (*the Cambridge and Nottingham Corpus of Discourse in English*, скраћено CANCODE). Испитујући семантичку склоност поменутих лексема, као и њихову семантичку прозодију, аутори истраживања закључују да на анализирани лексеме утиче контекст у ком се јављају (нпр. постоје разлике између њихових колоката и семантичке прозодије у интимној комуникацији, између чланова породице или партнера, у односу на комуникацију између професора и студената).

2. Неколико примера оквира из домена ризика

Као што је поменуто и на крају т. 1.3 овог рада, Филмор и Еткинсова разматрали су ограничења која су лексичкој анализи наметнута у традиционалном приступу лексикографији и у папирним издањима речника (Fillmore and B. T. Atkins 1992, 100–101), (Fillmore and S. Atkins 1994, 350–363). Пошто су упоредили анализе глагола *ризиковати* и именице *ризик* у референтним једнојезичним речницима с налазима о овим лексемама до којих су дошли анализом корпуса, схватили су да

16. Анализа је указала на то да се *одредба* јавља у чак 12 различитих оквира (Wasserscheidt and Hrštic 2020, 108).

ниједан од прегледаних речника не пружа задовољавајућу анализу и не бележи многе податке које су уочили прегледом примера из корпуса. Закључак је био да папирни речник, у ком је приказ линеаран, не може да представи тако сложен опис какав би био потребан да би се представили сви подаци важни за употребу неке речи. Због тога су осмислили онлајн речник заснован на оквирима, уместо папирног речника заснованог на лексемама, у ком је могуће представити резултате овакве анализе.

Овако замишљен онлајн речник омогућава да се прикажу поједини елементи оквира и њихове различите синтаксичке реализације, дакле све оно што спада у опис валенце (коју смо поменули у т. 1.4), као и односи између различитих оквира.

Алат за визуелизацију односа између оквира и елемената оквира (*FrameGrapher*)¹⁷ пружа могућност да се изабере почетни оквир и систематски истраже везе између оквира. Сlike 1–4 у наставку текста генерисане су управо коришћењем алата *FrameGrapher*.

2.1 Оквир *Ризична_ситуација*

У наставку је представљен оквир *Ризична_ситуација*.¹⁸ Након дефиниције дају се примери, а потом централни и периферни елементи оквира. Као што је поменуто, сваком елементу оквира додељује се посебна боја, тако да се и у тексту дефиниције елементи, ако се појаве, боје одговарајућим бојама. Лексичке јединице које призивају оквир *Ризична_ситуација* су: *опасност.n*, *опасан.a*, *ризик.n*, *рискантно.adv*, *ризичан.a*, *безбедан.a*, *безбедно.adv*, *небезбедан.a*, *претња.n*, *шкодљив/несигуран.a* (*n* је скраћеница за именице, *a* за придеве, *adv* за прилоге, *v* за глаголе). ЛЈ које призивају оквир мармирају се црном позадином у аотираном тексту. За сваки елемент оквира наводи се дефиниција и један обележен пример употребе.

17. *FrameGrapher*

18. За потребе овог рада превели смо опис неколико оквира и њихових елемената, уз напомену да је реч о опису везаном за енглески, са примерима из корпуса енглеског језика, прилагођеним српском језику ради илустрације принципа. Надамо се да ћемо опис оквира заснован на подацима из корпуса српског језика представити ускоро.

Ризична ситуација

Дефиниција:

Одређена **Ситуација** може (али не мора) да доведе до штетног догађаја који би задесио неку **Вредност**. Та **Ситуација** може бити неко стање, активност или неки кључни ентитет који мора бити схваћен као део неке шире **Ситуације**, која укључује и тај ентитет и **Вредност**. Иако је за разумевање овог оквира кључна идеја о штетном догађају, он не мора бити изражен као аргумент лексичких јединица у овом оквиру.

Да ли су **климатске промене** **ОПАСНЕ по човечанство**?

Купци се могу жалити на **ШКОЉИВЕ** **презивале**.

Највећа **ОПАСНОСТ** прети **нашој инфраструктури**.

Елементи оквира

Централни:

Вредност (Asset) [ass] Нешто што се сматра вредним и пожељним, а постоји могућност да ће му штета бити нанета или да ће бити изгубљено.

Мед није **БЕЗБЕДАН** **за бебе**.

Опасан ентитет (dangerous entity) Конкретан или апстрактан ентитет који може нанети штету **Вредности** или довести до њеног губитка.

Искључује: Ситуацију **Том човеку** је **РИЗИК** друго име!

Ситуација (Situation) [sit] **Ситуација** може довести до неког штетног догађаја. Већина **води** се слаже да није **БЕЗБЕДНО** **возити брже од 120 km/h**.

Периферни:

Околности (Circumstances) [i] **Околности** под којима је **Вредност** угрожена.

Степен (Degree) [deg] Одређба која изражава одступање тренутног нивоа безбедности од очекиване вредности, узимајући у обзир **Ситуацију** и стање на које указује циљна ЛЈ.

Домен (Domain) [dom] Терористи су наша **највећа** **ПРЕЂА**.

Учесталост (Frequency) [f] **Домен** у коме је **Ситуација** безбедна.

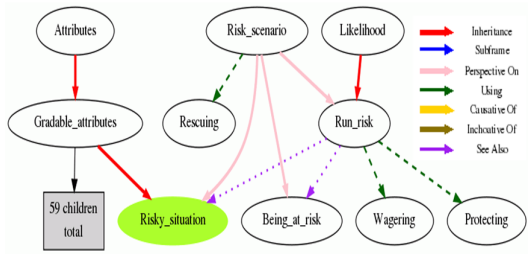
Место (Place) [pla] Све наше сајтове је **БЕЗБЕДНО** користити у **образовању**.

Време (Time) [tin] Колико често **Вредности** **долази** у **Ризичну ситуацију**.

Одређена локација на којој је **Ситуација** безбедна.

Често се може закључити да карактеристике неке локације чине одређене **Ситуације** безбедним или небезбедним.

Временски период током ког одређена **Ситуација** има прецизирани ниво сигурности.



Слика 1. Графички приказ односа између оквира *Ризична_ ситуација* (енгл. *Risky_ situation*) и с њим повезаних оквира

2.2 Оквир *Бити_угрожен*

Оквир *Бити_угрожен* (енгл. *Being_at_risk*) призивају следеће лексичке јединице: *опасност.п.*, *несигуран.а.*, *ризик.п.*, *безбедан.а.*, *сигуран.а.*, *безбедност.п.*, *поуздан.а.*, *рањивост.п.*, *рањив.а.* Боје одговарајућих елемената оквира исте су као у претходном примеру; овај оквир садржи и додатни елемент – *Штетни_догађај* (*Harmful_event*).

Бити_угрожен

Дефиниција:

Вредност је у неком стању у ком је изложена или подложна дејству **Штетног догађаја**, који може бити метонимијски призван дејством **Власног ентитета**. Речи које означавају релативну сигурност (одсуство ризика) такође су део овог оквира.

Нема делова које је **ЗАШТИЋЕНО** од искушења да уради оно што ради и његови пршњаци.

Наша држава ужасно греши покушавајући да се **ЗАШТИТИ** **од луци** – она мора да штити луце.

Уколико радим као багериста, **ми** си под **РИЗИКОМ** од губитка слуха због **изложености буци** доком радим.

Ви нисте **СИГУРНИ** од крађе података уколико немате заштиту од прислушкивања.

Елементи оквира:

Централни:

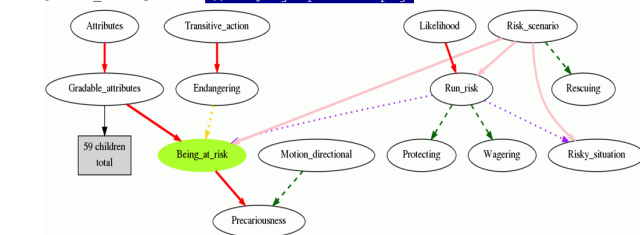
Вредност (Asset) [ass] Нешто што се сматра пожељним или драгоценим и што може бити изгубљено или оштећено.

Опасан ентитет (Dangerous entity) Закључани катанац гарантује да су **Информације СИГУРНЕ**. Конкретан или апстрактан ентитет који може да узрокује губитак или оштећење **Вредности** због њеног учешћа у **Штетном догађају**.

Штетан догађај (Harmful event) [har] Старомо се да ваш **ВМ** буде **БЕЗБЕДАН/ЗАШТИЋЕН** од **проваљника**. Догађај који се може одиграти или стање које се може одржати и које може довести до губитка или оштећења **Вредности**.

Искључује:

Опасан ентитет (Dangerous entity) Наш систем обезбеђује да информације које се чувају на хардверу буду **ЗАШТИЋЕНЕ** од напада хакера, као и **од покушаја физичке крађе**.



Слика 2. Семантички оквир *Бити_угрожен* (енгл. *Being_at_risk*)

2.3 Оквир *Изложити се_ризик*

Оквир *Изложити се_ризик* (енгл. *Run_risk*) призивају следеће лексичке јединице: *угрожен.а.*, *опасност.п.*, *ризик.п.*, *ризиковати.в.*, *угрозити.в.* На слици 3 дати су дефиниција, примери и елементи овог оквира.

Изложити се ризику

Дефиниција:

Протагониста је у потенцијално опасној ситуацији која може да се оконча **лошим исходом** по њега или њу. Опасност од губитка **Вредности** може да представља **Лош исход**. Не постоје назнаке да се **Протагониста** намерно излаже ризицијој ситуацији. Могуће је да **Протагониста** покушава да оствари неки **Циљ**, чиме доспева у опасну ситуацију. **Озбиљност** ризика којем се излаже такође може бити изражена. Постојао је **РИЗИК** да се све запали.

Елементи оквира:

Централни:

Рађа (Action) [Act] Рађа која изазива ризик.
Имплементација овог програма излаже нас **РИЗИКУ** да увредимо своје најверније бираче.
Нешто пожељно што Протагониста поседује или је с њим непосредно повезано и што може бити изгубљено или оштећено.
То је био велики РИЗИК по његову репутацију.

Вредност (Asset) [Asset]

Искључује:

Лош исход

(Bad outcome)

Лош исход (Bad outcome)

Bad]

Протагониста (Protagonist)

(Protagonist)

Периферни:

Бенефицијар

(Beneficiary) [ben]

Циљ

(Purpose) [Pur]

Озбиљност

(Severity) [Sev]

Ситуација коју би **Протагониста** хтео да избегне. **РИЗИКОВАО** је да изгуби **своје здравље**.

Особа којој прети неки **Лош исход**.

Протагониста има за циљ да његови поступци буду на корист **Бенефицијару**.

Све би **РИЗИКОВАЛИ** за своје најближе пријатеље.

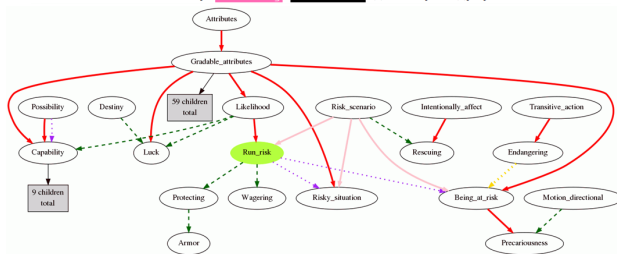
Жељена рађа или циљ за који **Протагониста** верује да ће га остварити.

Сви су они **РИЗИКОВАЛИ** да буду ухапшени **само да би се домогли Америке**.

Вероватноћа да ће се нешто лоше десити

Протагонисти.

Када се више не разликује добро и зло, човек је у **озбиљној ОПАСНОСТИ** да изгуби душу.



Слика 3. Семантички оквир *Изложити се ризику* (енгл. *Run_risk*)

Сценарио ризика

Дефиниција:

Вредност је у **ситуацији** за коју је вероватно да води до неког **штетног догађаја**, који ће лоше утицати на **Вредност**.

Елементи оквира:

Централни:

Вредност (Asset) [Ass] **Место** што се сматра пожељним или вредним и што може бити изгубљено или оштећено.

Штетни догађај (Harmful event) [Har] **Догађај** који може да се одигра или стање које може да потраје и које може довести до губитка или оштећења **Вредности**.

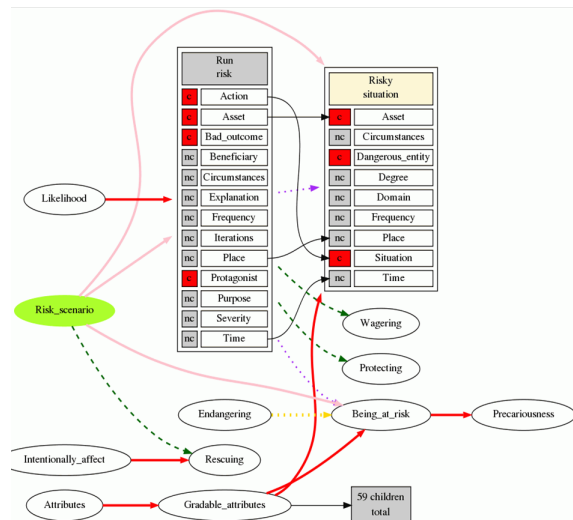
Ситуација (Situation) [Sit] **Ситуација** у којој **Вредност** није безбедна или заштићена.

Периферни:

Степен (Degree) [Deg] Одредба која изражава одступање од актуелног нивоа безбедности за **Вредност**, **ситуација** и стање означени самом циљном ЈЈ.

Место (Place) [Pla] **Место** за које важи одређени ниво безбедности.

Време (Time) [Tim] **Време** током ког важи одређени ниво безбедности.



Слика 4. Семантички оквир *Сценарио ризика* (енгл. *Risk_scenario*)

2.4 Сценарио_ризика

Слика 4 представља оквир *Сценарио_ризика* (енгл. *Risk_scenario*), са детаљним приказом оквира *Изложити се_ризика* (*Run_risk*) и *Ризична_ситуација* (*Risky_situation*), за које су обележени централни (енгл. *core*, скр. *c*) и периферни (енгл. *non-core*, скр. *nc*) елементи. На десној страни слике дата је легенда с врстама релација које се успостављају између оквира, на пример: наслеђивање, угао гледања, коришћење (приказане су и неке од релација које нисмо помињали: подоквир, узрочник и др).

3. NLTK омотач за Фрејмнет

NLTK (енгл. *Natural Language Toolkit*) је платформа за развој програма *Python* за обраду текста на природним језицима, која је једноставна за коришћење, а садржи бројне корпусе и лексичке ресурсе чији се број континуално увећава. NLTK омогућава различите врсте обраде текста, међу којима су: класификација, токенизација (енгл. *tokenization*), стемовање (енгл. *stemming*), тагирање (енгл. *tagging*), парсирање (енгл. *parsing*) и семантичко закључивање (енгл. *semantic reasoning*). У оквиру NLTK система имплементирани су и омотачи (енгл. *wrappers*) за друге програмске библиотеке за обраду природних језика, као и за важне лексичке ресурсе. Један од расположивих ресурса у оквиру NLTK је и Фрејмнет, који прати програмска библиотека за претраживање овог ресурса и екстракцију информација из његових оквира.

Као што је поменуто у Уводу (одељак 1.1 овог рада), оквир је појмовна структура¹⁹ која описује одређену врсту ситуације, ентитета или односа заједно са учесницима тих ситуација и односа. Физичка организација Фрејмнета у оквиру NLTK дистрибуције представља колекције XML (*Extensible Markup Language*) датотека разврстаних по каталозима: *frame*, *fulltext*, *lu*, *miscXML*, којима се приступа помоћу функција библиотеке или се могу директно претраживати и приказивати коришћењем XML докумената са XSL (енгл. *eXtensible Stylesheet Language*) трансформацијама: *frameIndex*, *luIndex*, *fulltextIndex*. У овом одељку ћемо приказати примере коришћења функција омотача Фрејмнета.

19. [FrameNet](#) и [NLTK](#)

Да би се добила листа свих оквира у Фрејмнету, може се користити функција `frames()`. Следећи код приказује иницијализацију рада са корпусом Фрејмнет и показује да верзија која је публикована уз NLTK има 1221 оквир.

```
from nltk.corpus import framenet as fn
len(fn.frames())
```

Да бисмо пронашли оквире који у називу садрже реч 'risk', користи се опција:

```
fn.frames(r'risk')
```

која као резултат даје:

```
[<frame ID=1560 name=Being_at_risk>,
 <frame ID=378 name=Run_risk>].
```

Имајући у виду осетљивост на мала и велика слова, за упит:

```
fn.frames(r'Risk')
```

добија се другачији резултат:

```
[<frame ID=1763 name=Risk_scenario>,
 <frame ID=1762 name=Risky_situation>]
```

Ако се зада регуларни израз '(?i)risk' функцији `frame()`, добија се листа од сва четири поменута оквира (т. 2.1–2.4), чија се имена подударају с прослеђеним обрасцем, јер '(?i)' указује на то да се не прави разлика између малих и великих слова.

Детаљи одређеног оквира могу се добити функцијом `frame()`, којој се проследи као параметар број оквира, рецимо `f=fn.frame(1762)`, што ће вратити податке за оквир Ризична_ситуација (*Risky_situation*).²⁰

Уз то, може се прићи деловима оквира, рецимо називу оквира (`f.name`), његовој дефиницији (`f.definition`), елементима оквира (`f.FE`), лексичким јединицама (`f.lexUnit`), релацијама између оквира (`f.frameRelations`), што је илустровано следећим кодом:

```
f = fn.frame('Risky_situation')
print(sorted([e for e in f.FE]))
print([r for r in f.frameRelations])
```

20. Подаци за оквир *Ризична_ситуација*

који даје следећи резултат:

```
[ 'Asset', 'Circumstances', 'Dangerous_entity', 'Degree', 'Domain',
  'Frequency', 'Place', 'Situation', 'Time' ]
[ <Parent=Gradable_attributes - Inheritance →
  Child=Risky_situation>,
  <MainEntry=Run_risk - See_also →
  ReferringEntry=Risky_situation>,
  <Source=Run_risk - ReFraming_Mapping →
  Target=Risky_situation>,
  <Neutral=Risk_scenario - Perspective_on →
  Perspectivized=Risky_situation> ]
```

4. Анализа речи *ризик* на корпусу рударства

Развој једнојезичног корпуса из домена рударства почео је као део пројекта везаног за управљање рударском пројектном документацијом коришћењем језичких технологија (Томашевић et al. 2018, 996). Тада је једнојезични корпус обухватао текстове из рударског домена и сродних истраживања, са укупно 172 документа (на српском језику) и 2,7 милиона речи у првом издању (997). Током даљих истраживања (Kitanović 2021) проширен је са додатна 63 документа. Тренутна верзија садржи 4,1 милион речи. Класификован према изворима, корпус обухвата пројектну документацију (26%), законодавство (11%), докторске дисертације (31%), уџбенике и осталу рударску литературу (32%) (Kitanović et al. 2021, 8).

Резултат претраге у CQL-у²¹ (енгл. *Corpus Query Language*) анализира се на различите начине: листа фреквенција, колокације, конкорданце са ужим и ширим контекстом. Слика 5 приказује конкорданце, добијене из веб-апликације Лексимирика²² за управљање електронским речницима (Stanković et al. 2018), за образац придев-именица за именицу *ризик*, док се на слици 6 може видети хистограм са фреквенцијама препознатих облика за исти образац, екстрахованих из рударског корпуса, доступног на корпусној платформи отвореног кода *NoSketch Engine* (Kilgariff et al. 2004).²³ Инстанцу на локалном серверу одржавају чланови Друштва за језичке ресурсе и

21. [Corpus Querying](#)

22. [Лексимирика](#)

23. [NoSketch на JePTexy](#), [NoSketch Engine](#)

Concordances – Profil 1 – Microsoft Edge
 https://leximirkaj.rerth.rs/CCP/Noske

A(N)

na najmanju moguću meru, odnosno otklanjanje	profesionalnih rizika	. Strategija teži da se u ovom periodu broj
na najmanju moguću meru, odnosno otklanjanje	profesionalnih rizika	. Strategija teži da se u ovom periodu broj
inspektora rada sa novim tehnologijama i	novim rizicima	. savremenim pristupima i praksama u oblasti
i zdravlja na radu uzimajući u obzir	posebne rizike	koji se pojavljuju u određenim delatnostima . o
uzajamna povezanost, što samo još povećava	potencijalne rizike	za ukupnu realizaciju procesa rekultivacije .
velikih količina otpada po konkurentnoj ceni a	niskom riziku	po životnu sredinu ; 2. ekonomski isplativ
, potencijalno moguće ozbiljnije povrede,	mali rizik	fatalnog kraja, gubici radnog vremena Nizak
i normalna komunikacija Neophodna prva pomoć,	mali rizik	od ozbiljnih povreda Zanemarujući Nemerljivi
6, jer osim snabdevanja gasom, kod nje postoji i	izvesni rizik	od povraćaja investicije, što bi moglo
odgovora često veoma zahtevan, složen, i sa	prisutnim rizicima	. Konačno rešenje, kako smo već istakli u
i sl. • Prisustvo konfliktnih situacija,	povišenih rizika	i nepovoljnih događaja, npr. interakcija
система zaštite na radu : 1) radnim mestima sa	povećanim rizikom	; 2) zaposlenima raspoređenim na radna mesta
: 2) zaposlenima raspoređenim na radna mesta sa	povećanim rizikom	i lekarskim pregledima zaposlenih
može da ima previd pojedinih opasnosti. Psiho-	socijalni rizici	se obično prevede, kao i rizici u vezi sa
, a takođe je ostvaren napredak i u proceni	profesionalnih rizika	i sistematizaciji profesionalnih bolesti .
ili smanjenja rizika. Radno mesto sa	povećanim rizikom	jeste radno mesto utvrđeno aktom o proceni
je da se nekontrolisane opasnosti prevedu u	kontrolisani rizik	i da se na taj način bolje zaštite zaposleni i
identifikovanju i kontrole zdravstvenih i	sigurnosnih rizika	organizacije i eliminisanju ili smanjivanju
organizacije i eliminisanju ili smanjivanju	potencijalnog rizika	od nezgoda na prihvatljivi nivo, poštujući pri
najvišeg mogućeg nivoa bezbednosti i	minimalnog rizika	moraju se dokumentovati uključujući i zapise

Слика 5. Конкорданце обрасца придев–именица за именицу *ризик*

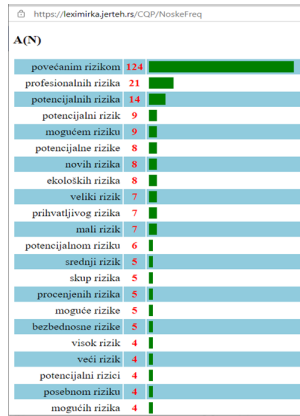
технологије JePTex.²⁴ За тагирање корпуса обучен је *Treegger* модел за српски (Krstev and Vitas 2005; Utvić 2011), (Stanković et al. 2020, 3957), коришћењем ручно анотираниог корпуса и Српских морфолошких речника (Krstev 2008).

Рударски корпус је публикован и на платформи *Sketch Engine*²⁵ (Kilgarriff et al. 2014), на којој се могу поставити различити упити. Рецимо, можемо извући конкорданце изабране леме или израза, потом колокације за лему, тезаурус сродних речи, затим такозвану скицу речи (енгл. *Word Sketch*) или видети разлику у употреби две сродне речи (енгл. *Word Sketch Difference*). Модул екстракције скице речи, који су развили Kilgarriff et al. (2004), помаже у изградњи Фрејмента и сличних ресурса, убрзава доношење одлука о издвајању различитих значења вишезначних лексема (Baker 2012, 274).

Скица речи даје брзи преглед понашања изабране лексеме тако што прикупља информације из хиљада и милиона примера њене употребе и пружа сажетак категоризованих колокација на једној страници, с везама до појединачних примера. Слика 7 приказује пример скице речи за именицу *ризик* – један поглед на дату страницу довољан је да се

24. JePTex

25. Sketch Engine



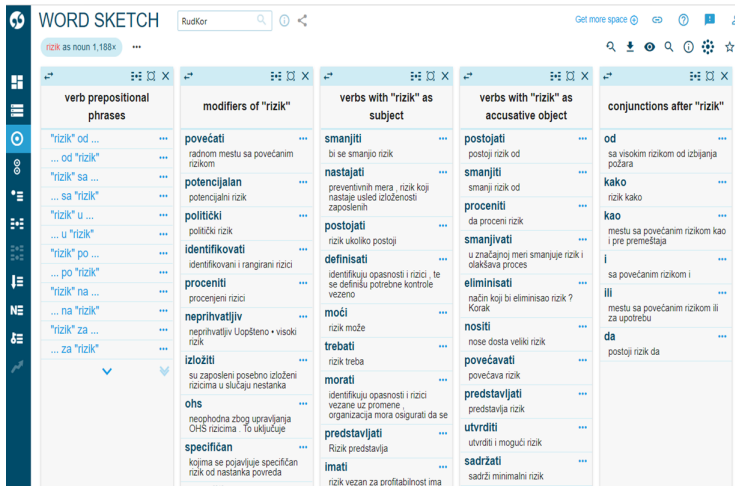
Слика 6. Хистограм са фреквенцијама облика обрасца придев-именица за именицу *ризик*

схвати како се та реч користи. Приказан је део скице речи: у првој колони видимо оно што се у српском језику назива предлошко-падежним конструкцијама (у енглеској граматичкој терминологији предлошким фразама):²⁶ *ризик од/са/у/по/на/за...*, за које је даље могуће доћи до конкорданци које одговарају конкретном обрасу кликом на "...". У другој колони представљени су модификатори речи *ризик*: трпни придев глагола (*повећан/идентификован/процењен/ ... ризик*) или придев (*потенцијалан/политички/неприхватљив/ ... ризик*). У трећој колони налазе се глаголи уз које се *ризик* јавља као субјекат, на пример: *смањити се/настајати/постојати/...* Даље следе обрасци у којима се *ризик* јавља као објекат уз глаголе: *смањити/проценити/ ...*

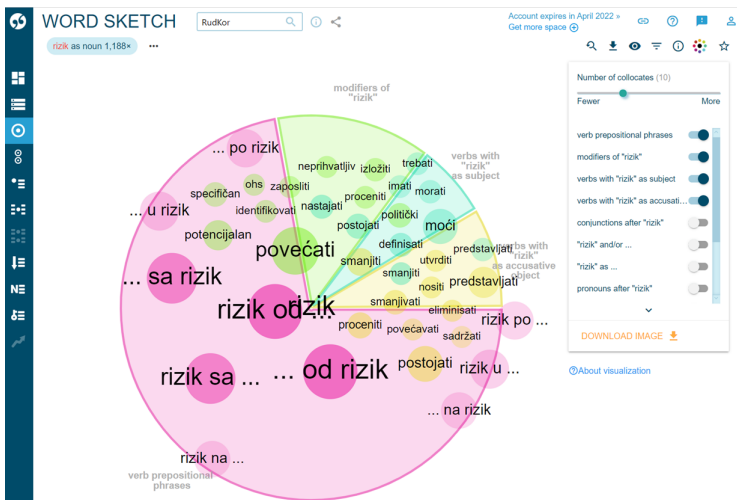
Приказ колокација у виду динамичког дијаграма дат је на слици 8. Може се видети да већину колоката чини образац предлошко-падежне конструкције (на слици: *verb prepositional phrases*). Десни део слике приказује могућности подешавања: који ће обрасци бити приказани и која је минимална фреквенција колокација да би биле укључене на граф.

Истраживање колокација од изузетног је значаја (на пример, у лексикографији, важно је навести најчешће колокате одређених ЛЈ; колокације су кључне и у учењу језика, али и у различитим задацима

26. Важно је напоменути да алати и аутоматско рангирање нису сасвим прилагођени за српски језик и да се јављају грешке, које се морају накнадно исправљати, али су свакако врло драгоцени.

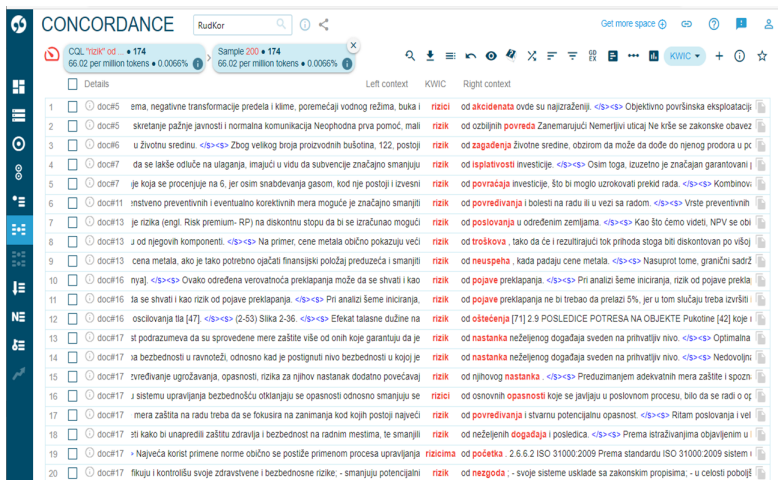


Слика 7. Скица речи *ризик* у алату *Sketch engine*



Слика 8. Графички приказ колокација за именицу *ризик* у алату *Sketch engine*

обраде природних језика). Полазећи од скице речи за облик *ризик* и колокацију *ризик од*, могу се приказати одговарајуће конкорданце (слика 9).

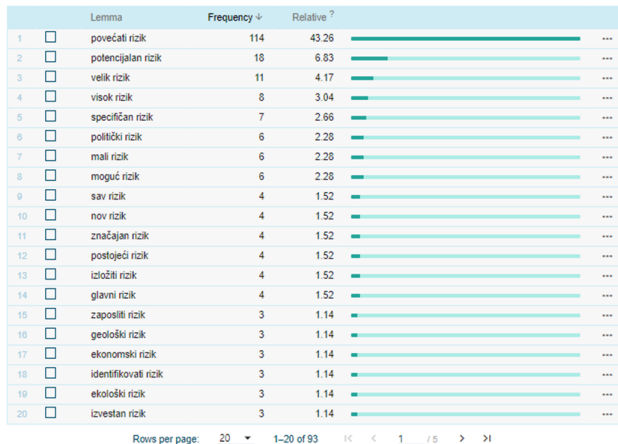


Слика 9. Конкорданце у алату *Sketch engine*

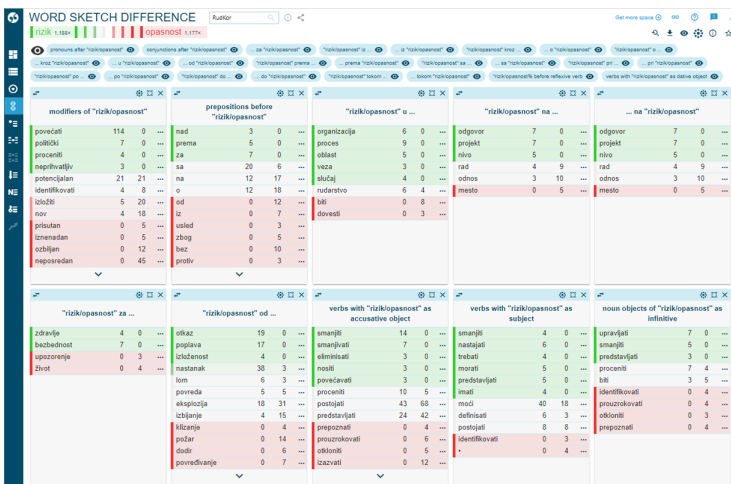
Скица речи даје брзи преглед за унапред припремљена правила, али се корисни резултати могу добити и CQL упитима. На пример, на питање (*Од*) чега је *ризик* (тј. *од чега постоји ризик*)?, одговор бисмо могли наћи следећим CQL упитом [lemma="ризик"] [tag="N"]. Одговор на питање *Какав је ризик*? дао би упит [tag="A"] [lemma="ризик"]; или, ако желимо да дозволимо размак између речи (не више од 5), онда бисмо написали упит на следећи начин: [tag="V"] [word!=" "] 0,5 [lemma="ризик"].

Осим што се може дати испис статистичких фреквенција, оне се могу приказати и визуелно, линијама на десној страни. Слика 10 приказује фреквенције појављивања именице *ризик* уз поједине придеве и глаголе.

Слика 11 приказује панел разлика у скицама речи (енгл. *Word Sketch Difference*). Панел разлика представља проширење модула скице речи. Генерише скице за две речи и упоређује их, што омогућава да се уоче разлике у њиховој употреби. Функција је посебно корисна за блискозначнице, за антониме и речи из истог семантичког поља. Са



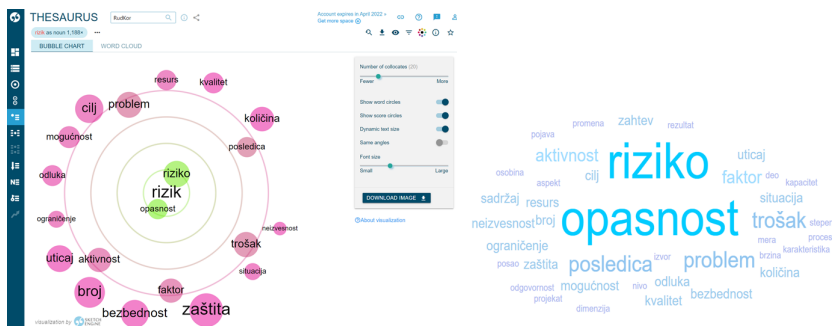
Слика 10. Фреквенције колокација за именицу *ризик* у алату *Sketch engine*



Слика 11. Скица разлика употребе речи *ризик* и *опасност*

слике 10 се, на пример, може видети да се уз *ризик* користе колокати *повећати*, *политички*, *проценити*, *прихватљив*, док се уз *опасност* јављају *непосредан*, *озбиљан*, *изненадан*.

Аутоматско генерисање тезауруса за задату реч проналази синониме или речи које припадају истој категорији (истом семантичком пољу) и приказује их у виду табеле, с могућношћу позивања одговарајућих конкорданци, скице речи, разлика, тезауруса. Могуће је и графичко приказивање тезауруса, као на слици 12, на којој се могу видети аутоматски генерисане речи које припадају истој категорији (семантичком пољу) као задата реч *ризик*, лево у виду графикона с мехурићима (енгл. *bubble graph*), а десно у виду облака речи. Листа речи тезауруса направљена је на основу контекста у којем се речи појављују у изабраном текстуалном корпусу, ослањајући се на теорију дистрибуционе семантике, која укратко каже да су речи које се појављују у истом контексту сличног значења. Да би се одредили синоними, упоређују се скице речи свих речи исте врсте, а оне које деле највећи део колоката наводе се као сличне речи. Оцена²⁷ дата за сваки синоним указује на проценат заједничких колоката.



Слика 12. Графички приказ тезауруса речи *ризик*

27. Статистичке формуле које се користе у алату *Sketch engine*: [статистика/формуле](#)

5. Закључак

У раду су приказани резултати прелиминарних истраживања у вези с могућностима примене теорије семантике оквира и принципа коришћених у изради семантичке мреже Фрејмнет на примерима из домена оквира ризика прилагођених српском језику. Уз то, представљена је платформа NLTK, погодна за коришћење разних језичких ресурса, као и систем *Sketch Engine* за корпусну анализу текста.

Показано је да анализа у оквиру Фрејмнета нуди детаљну, структурирану интерпретацију, која се даље може користити у различитим апликацијама за обраду природног језика, посебно у проналажењу и смисленом организовању текстова и у природнијој интеракцији између човека и рачунара, нарочито у све популарнијим дијалог-системима (енгл. *chatbot*). Систем дијалога који комуницира с корисницима мора бити у стању да препозна различите речи које призивају исту ситуацију или се односе на исти ентитет да би се извршило успешно препознавање намере.

Погодно је то што се енглески лексикон Фрејмнет може попунити лексичким подацима из других језика, рецимо српског (уз задржавање оних информација о неком оквиру које су заједничке за два језика и модификацију оних које су језички специфичне), тако да је модел погодан за вишејезичне ресурсе и апликације.

Представљена истраживања тек наговештају могућности прилагођавања Фрејмнета за српски језик и његовог поравнавања с таквим базама података за друге језике у будућности. Планирана даља истраживања треба да обухвате и могућности поравнања и заједничког коришћења српског Ворднета и Фрејмнета, као и њихово узајамно допуњавање. У том послу руководићемо се препорукама које дају Tonelli and Pighin (2009).

Овим истраживањима подстиче се и развој корпусне лексикографије српског језика, као и модернизација његовог граматичког и лексикографског описа. Тој модернизацији свакако би допринеле и студије случаја у којима би се поредиле обраде неких вишезначних лексема у Речнику САНУ с њиховим описом из угла теорије семантике оквира.

Могућности за будућа истраживања у овој области су велике. Примена теорије семантике оквира и методологије која се користи у изради Фрејмнета, као и представљених алата на српски језик захтеваће решавање бројних питања; на основу овог рада сматрамо да

ће бити веома изазовно сагледати уз помоћ појма нултог појављивања (*null-instantiation*) допуне транзитивних глагола које се не морају експлицирати, али се подразумевају (*кувати*, *писати* и сл.), као и начин на који се у описним речницима представљају такви случајеви. Уз то, сматрамо да би случајеве нултог појављивања (а у методологији Фрејмнета су препозната три таква случаја) било добро увести у српску граматику.

Захвалност

Овај рад је подржало Министарство просвете, науке и технолошког развоја Републике Србије према уговору бр. 451-03-9/2021-14, који је склопљен са Институтом за српски језик САНУ. Приступ алату SketchEngine је обезбеђен пројектом ELEXIS који се финансира из Horizon 2020 истраживачког и иновационог програма Европске уније, грант број 731015.

Литература

- Atkins, Beryl T. S. 1994. "Analyzing the verbs of seeing: a frame semantics approach to corpus lexicography." In *Annual Meeting of the Berkeley Linguistics Society*, 20:42–56. 1.
- Atkins, Sue, Charles J Fillmore, and Christopher R Johnson. 2003. "Lexicographic relevance: Selecting information from corpus evidence." *International Journal of Lexicography* 16 (3): 251–280.
- Baker, Collin F. 2012. "FrameNet, current collaborations and future goals." *Language Resources and Evaluation* 46 (2): 269–286.
- Boas, Hans C., and Ryan Dux. 2017. "From the past into the present: From case frames to semantic frames." *Linguistics Vanguard* 3 (1): 20160003. <https://doi.org/doi:10.1515/lingvan-2016-0003>.
- Brač, Ivana, and Ana Ostroški Anić. 2019. "From concept definitions to semantic role labeling in specialized knowledge resources." In *Proceedings of the 13th International Conference of the Asian Association for Lexicography*, 604–611.

- Fillmore, Charles J. 1976. "Frame semantics and the nature of language." In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, 280:20–32. 1. New York.
- Fillmore, Charles J. 1982. "Frame semantics." *Linguistic society of Korea (ed.), Linguistics in the morning calm*, 111–137.
- Fillmore, Charles J, and Beryl T Atkins. 1992. "Toward a frame-based lexicon: The semantics of RISK and its neighbors." *Frames, fields and contrasts: New essays in semantic and lexical organization* 75:102.
- Fillmore, Charles J, and Sue Atkins. 1994. "Starting where the Dictionaries Stop: The Challenge of Corpus Lexicography." In *Computational Approaches to the Lexicon*, edited by Sue Atkins and Antonio Zampolli, 349–393. Oxford: OUP.
- Fillmore, Charles J, Miriam RL Petruck, Josef Ruppenhofer, and Abby Wright. 2003. "FrameNet in action: The case of attaching." *International journal of lexicography* 16 (3): 297–332.
- Gantar, Polona, Kristina Štrkalj Despot, Simon Krek, and Nikola Ljubešić. 2018. "Towards semantic role labeling in Slovene and Croatian." In *Proceedings Conference on Language Technologies and Digital Humanities in Ljubljana*, 93–98.
- Gildea, Daniel, and Daniel Jurafsky. 2002. "Automatic labeling of semantic roles." *Computational linguistics* 28 (3): 245–288.
- Hamilton, Craig, Svenja Adolphs, and Brigitte Nerlich. 2007. "The meanings of 'risk': A view from corpus linguistics." *Discourse & Society* 18 (2): 163–181.
- Jurafsky, Dan, and James H Martin. 2020. "Semantic Role Labeling and Argument Structure." Chap. 19 in *Speech and Language Processing*, 3rd ed. December 30, 2020 draft.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychl, and Vít Suchomel. 2014. "The Sketch Engine: ten years on." *Lexicography* 1 (1): 7–36.
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. "Itri-04-08 the sketch engine." *Information Technology* 105 (116).

- Kitanović, Olivera. 2021. “Ontološki model upravljanja rizikom u rudarstvu.” PhD diss., Univerzitet u Beogradu, Rudarsko-geološki fakultet. <https://uvidok.rcub.bg.ac.rs/bitstream/handle/123456789/4305/Dokorat.pdf?sequence=1>.
- Kitanović, Olivera, Ranka Stanković, Aleksandra Tomašević, Mihailo Škorić, Ivan Babić, and Ljiljana Kolonja. 2021. “A Data Driven Approach for Raw Material Terminology.” *Applied Sciences* 11 (7): 2892.
- Krstev, Cvetana. 2008. *Processing of Serbian. Automata, texts and electronic dictionaries*. Faculty of Philology of the University of Belgrade.
- Krstev, Cvetana, and Duško Vitas. 2005. “Corpus and Lexicon-Mutual Incompleteness.” In *Proceedings of the Corpus Linguistics Conference*, 14:17.
- Pradhan, Sameer, Wayne Ward, Kadri Hacıoglu, James H Martin, and Dan Jurafsky. 2005. “Semantic role labeling using different syntactic views.” In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)*, 581–588.
- Rasulić, Katarina, and Duška Klikovac. 2014. *Jezik i saznanje: hrestomatija iz kognitivne lingvistike*. Univerzitet u Beogradu, Filološki fakultet.
- Ruppenhofer, Josef, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Revised November 1, 2016. https://framenet.icsi.berkeley.edu/fndrupal/the_book.
- Stanković, Ranka, Cvetana Krstev, Biljana Lazić, and Mihailo Škorić. 2018. “Electronic dictionaries-from file system to lemon based lexical database.” In *Proceedings of the 11th LREC - W23 6th Workshop on Linked Data in Linguistics: Towards Linguistic Data Science (LDL-2018), LREC 2018, Miyazaki, Japan, May 7-12, 2018*, 48–56.
- Stanković, Ranka, Branislava Šandrih, Cvetana Krstev, Miloš Utvić, and Mihailo Skoric. 2020. “Machine Learning and Deep Neural Network-Based Lemmatization and Morphosyntactic Tagging for Serbian.” In *Proceedings of The 12th LREC – Language Resources and Evaluation Conference*, 3954–3962.

- Tomašević, Aleksandra, Ranka Stanković, Miloš Utvić, Ivan Obradović, and Božo Kolonja. 2018. "Managing mining project documentation using human language technology." *The Electronic Library*, <https://doi.org/10.1108/EL-11-2017-0239>.
- Tonelli, Sara, and Daniele Pighin. 2009. "New features for framenet-wordnet mapping." In *Proceedings of the thirteenth conference on computational natural language learning (CoNLL-2009)*, 219–227.
- Utvić, Miloš. 2011. "Annotating the Corpus of Contemporary Serbian." *IN-FOtheca* 12, no. 2 (December): 36a–47a.
- Wasserscheidt, Philipp, and Andrea Hrstić. 2020. "Legal Variation? A Frame Analysis of Croatian and Serbian in the Domain of Law." *Mediterranean Language Review* 27:87–112.
- Драгићевић, Рајна. 2007. *Лексикологија српског језика*. Београд: Завод за уџбенике.
- Марковић, Александра. 2017. "Однос граматике и речника – граматика инхерентна описним речницима српског језика." *Наш језик XLVIII* (1-2): 27–43.
- Поповић, Љубомир. 2003. "Интегрални речнички модели и њихов значај за лингвистички опис и анализу корпуса." *Научни састанак слависта у Вукове дане* 31 (1): 201–220.

