



UNIDIVE WG2:

MWE dictionaries as Linked Data & Parallel corpus
as Linked Data, Use Case on T2.2 corpus

prof. dr Ranka Stanković

UniDive 3rd general meeting, Tuesday 28 January 2025: WG2 meeting

Overview

- Continuation of work presented in
 - Stanković, Ranka, Christian Chiarcos, and Milica Ikonić Nešić. "Leveraging Linked Data, NIF, and CONLL-U for Enhanced Annotation in Sentence Aligned Parallel Corpora." UniDive 2nd general meeting, Naples. 2024. [abstract](#), [poster](#)
- Transformation of sentences from TSV to NIF
 - Jupiter notebook
 - ttl analysis
 - Fuseki endpoint
- NEL (Wikidata)
 - Wikidata SPARQL, geoJSON, QGIS
- VMWE & MWE
 - INCEPTION, Leximirka, Ontolex-lemon,...
 - Open questions?

Introduction

- Linked Data (LD) for the lexicon-corpus interface
 - interlinking MWE lexicon entries with their occurrences in corpora
- Publishing of aligned and annotated corpus as LD
 - employing NLP Interchange Format (NIF) and CoNLL-RDF
- UniDive T2.2: Design of a lexicon-corpus interface
 - to extend the ELEXIS-WSD Parallel Sense-Annotated Corpus
 - not in N^o sentences (2,024) and repository, but on languages, layers, enhanced technology.
- Apart from new languages and new annotation layers,
 - also corpus annotation should be upgraded to allow for interlinking MWE lexicon entries with their occurrences in corpora (as LD).

MWE dictionaries as Linked Data

- Modeling alternatives:
 - OntoLex, <https://www.w3.org/2016/05/ontolex>
 - Data Model for Lexicography (DMLex)
<https://docs.oasis-open.org/lexidma/dmlex/v1.0/dmlex-v1.0.html>
- Ontolex
 - Core module Lemon (general data structures),
 - Decomp (for the internal structure and combinatory semantics of MWEs),
 - Morph module (MWE morphology),
 - Lexicog module for lexicography,
 - FRaC (Frequency, Attestations, and Corpus-based Information)

```

:le_blood_pressure a ontolex:LexicalEntry,
                    ontolex:MultiwordExpression;
  ontolex:canonicalForm [ontolex:writtenRep "blood pressure"@en];
  lexinfo:partOfSpeech lexinfo:noun;
  ontolex:sense [ontolex:reference
    <https://dbpedia.org/page/Blood\_pressure>];
  decomp:constituent :cm_blood;
  decomp:constituent :cm_pressure;
  rdf:_1 :le_blood; # lexical
  rdf:_2 :le_pressure. # entries

```

component of cannonical form

```

:cm_blood a decomp:Component;
  decomp:correspondsTo :le_blood.
...
:le_krvni_tlak a ontolex:LexicalEntry, ontolex:MultiwordExpression;
  ontolex:canonicalForm [ontolex:writtenRep "krvni tlak"@sl];
...

```

```

:tranSetEN-SL a vartrans:TranslationSet ;
  dc:source <http://hdl.handle.net/...> ;
...

```

simplified naming

```

:tranSetEN-SL vartrans:trans blood_pressure-ensns-krvni_tlak-slsns .
:blood_pressure-ensns a ontolex:LexicalSense ;
  ontolex:isSenseOf :le_blood_pressure .
:krvni_tlak-slsns a ontolex:LexicalSense ;
  ontolex:isSenseOf :le_krvni_tlak .
:blood_pressure-ensns-krvni_tlak-sns-trans a vartrans:Translation ;
  vartrans:source :blood_pressure-ensns ;
  vartrans:target :krvni_tlak-slsns .

```

MWE lexical entry example:

- blood pressure
- with translations and links between senses and DBpedia entries.

Parallel corpus as Linked Data

- Parallel corpora as LD benefits of increased accessibility, interoperability, semantic enrichment, community collaboration, and the promotion of open science.
- Through practical implementations and experiments, we demonstrate the efficiency of incorporating LD principles, NIF, and CoNLL-RDF in aligned parallel corpora annotation.
- We contribute to the ongoing discourse on leveraging LD in the realm of aligned parallel corpora annotation, paving the way for more robust and efficient NLP applications.

Parallel corpus as Linked Data

```
·<http://url>·a·nif:ContextCollection·;  
···nif:hasContext·<http://url/enwsd>···.
```

```
<http://url/enwsd>·a·nif:Context,  
·····nif:OffsetBasedString·;  
·····nif:beginIndex·"0"···;  
·····nif:endIndex·"49"···;  
·····nif:isString·"He·is·named·after·  
·····the·astronomer·Galileo·Galilei."···.
```

```
<http://url/enwsd#offset_0_49_0>·a·  
·nif:OffsetBasedString,·nif:Phrase·;  
·nif:anchorOf·"Galileo·Galilei."·;  
·nif:beginIndex·"33";  
·nif:endIndex·"49"·;  
·nif:referenceContext·<http://url/enwsd>·;  
·nif:taMsClassRef/itsrdf:taIdentRef·wd:Q307;··  
·itsrdf:taClassRef·dbo:Person,wd:Q5,  
···<http://nerd.eurecom.fr/ontology#Person>.
```

- For establishing links between sentences that are translation equivalents, **skos:closeMatch** from SKOS (Simple Knowledge Organization System) can be used
- (indicates that two objects are sufficiently similar that they can be used alternately in applications).

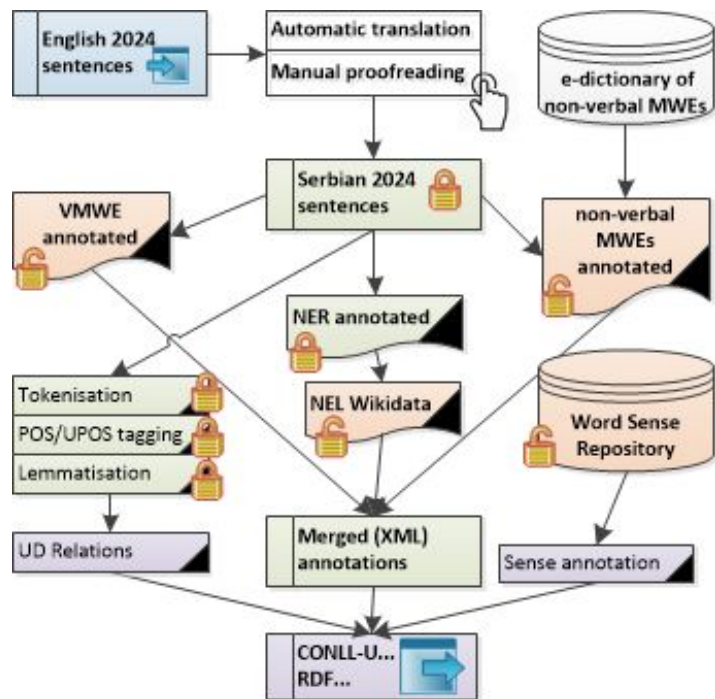
Lexicon-corpus interface

```
# multiword inflected form attestation
:le_blood_pressure frac:attestation [
  frac:quotation "Physical examination may
  sometimes reveal low blood pressure,
  high heart rate, or low oxygen saturation."@en;
  frac:observedIn :EWSO].

:le_krvni_tlak frac:attestation [
  frac:quotation "Telesni pregled včasih
  razkrije nizek krvni tlak, visok srčni
  utrip ali nizko nasičenost s kisikom."@sl;
  frac:observedIn :EWSO].
```

Ontolex-FRaC

T2.2 Case Study on Serbian



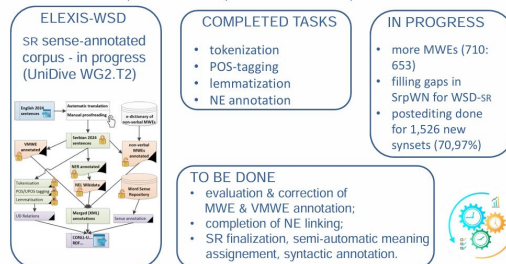
Merged annotation in
INCEPTION

<http://inception.jerteh.rs/> :

- POS
- Lemma
- NER (8 categories)
- NEL
- VMWE +MWE

Progress in SR-ELEXIS Semantic Annotation: Focusing on Multiword Expressions, Named Entities, and Sense Repository

Cvetana Krstev, Ranka Stanković, Aleksandra Marković, Milica Ikončić Nešić



Similar to other languages

Named entity
"Prvi svetski rat"
 ID: 5317
 identifier: <http://www.wikidata.org/entity/Q361>

Label: World War I
Description: 1914–1918 global war, centred in Europe, between the Allied and Central Powers

World War I | EVENT
 1 | svetski rat | NOUN
 ADJ | A
 prvi
 NOUN | N
 rat
 VERB | V
 izbio
 AUX | V
 jesam
 ADJ | A
 1914.

12 Prvi svetski rat izbio je 1914.

1 | LVCause
 NOUN | N
 većina
 NOUN | N
 stanovništvo
 1 | činiti | LVCause
 VERB | V
 čine
 NOUN | N
 katolički
 PUNCT | SENT

13 Večinu stanovništva čine katolici

NOUN | N
 član
 DET | PRO
 svaki
 ADJ | A
 političke
 NOUN | N
 grupa
 1 | IRV
 PART | PAR
 se
 1 | sastajati se | IRV
 VERB | V
 sastaju
 ADP | PREP
 pre
 ADJ | A
 velik
 NOUN | N
 sastanaka
 SCONJ | CONJ
 kako
 AUX | V
 bi
 VERB | V
 usvojili
 ADJ | A
 zajedničke
 NOUN | N
 stavove
 PUNCT | SENT

14 Članovi svake političke grupe se sastaju pre velikih sastanaka kako bi usvojili zajedničke stavove

WIKIDATA

Prvi svetski rat (Q361)

svetski rat vođen između 1914. i 1918.

Jezik	Naziv	Opis	Povezanimi
srpski / srpski	Prvi svetski rat	svetski rat vođen između 1914. i 1918.	1. svetski rat I svetski rat Prvi svjetski rat Veliki rat
engleski	World War I	global war originating in Europe, 1914–1918	World War One WW1 World War 1 World War the First World War 1st World War the 1st World War WW I The Great War WW 1 World War, 1914-1918 Great War WWI The War to End All Wars First World War
engleski (Sjedinjene Američke Države)	Oznaka nije definisana	Opis nije definisana	
slovenački	prva svetovna vojna	vojna med letoma 1914 in 1918, v kateri je sodelovalo približno stideset držav in v kateri je umrlo približno 17 milijonov ljudi	velika vojna 1. svetovna vojna

#Text=Prvi svetski rat izbio je 1914.

12-1	1184-1188	Prvi	A	ADJ	http://www.wikidata.org/entity/Q361	[17]	EVENT	[17]
12-2	1189-1196	svetski	A	ADJ	http://www.wikidata.org/entity/Q361	[17]	EVENT	[17]
12-3	1197-1200	rat	N	NOUN	http://www.wikidata.org/entity/Q361	[17]	EVENT	[17]
12-4	1201-1206	izbio	V	VERB				
12-5	1207-1209	je	V	AUX				
12-6	1210-1215	1914.	A	ADJ				

Similar to other languages

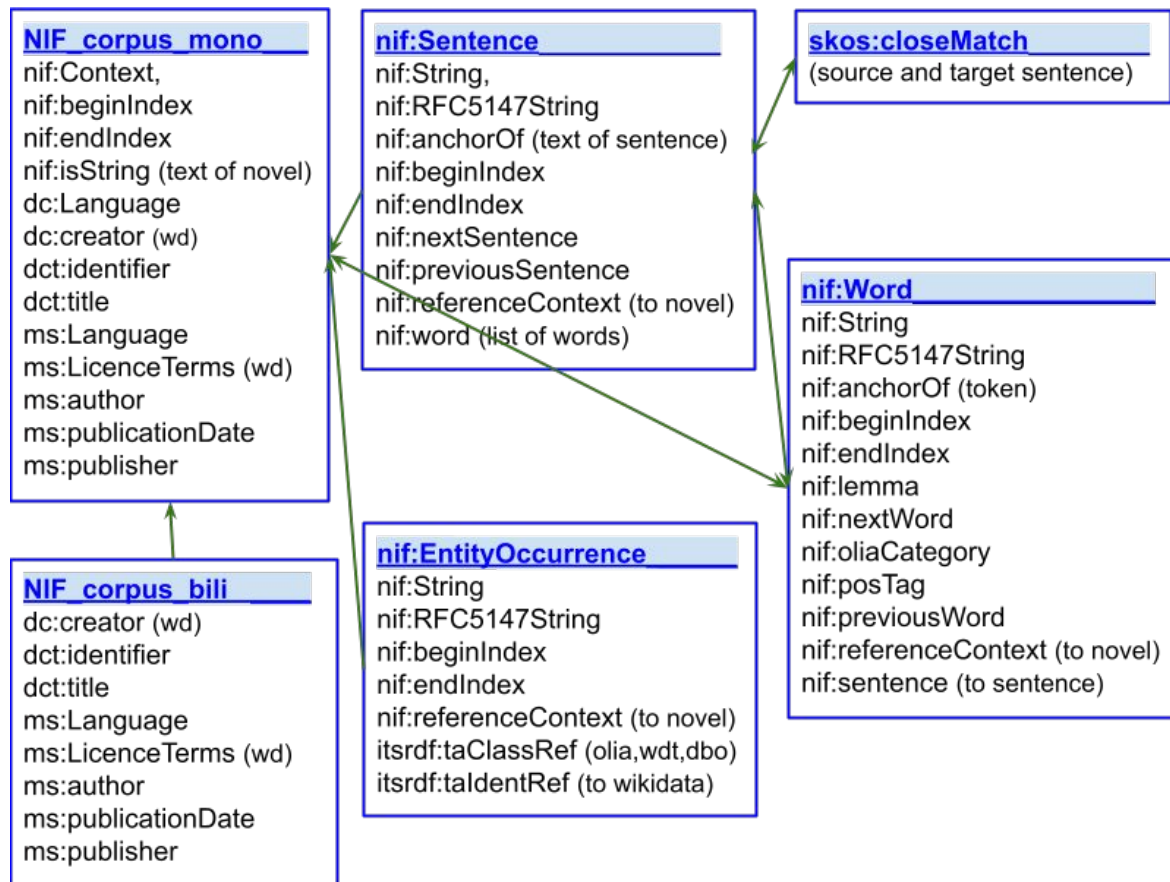
cvetana.krstev: SR-ELEXSIS-WSD/sr-elexsis-mwe_vmwe.tsv

1893	U početku	Bog	stvari nebo i zemlju.
1894	Kontrola temperature	se	održavala kroz toplotnu izolaciju i s
1895	Agrarna politika	i tehnologija u Sjedinjenim Državama i	Evropi dovele su do nižih cena hra
1896	U razvijenim zemljama nivoi gojaznosti odraslih i procenat tinejdžera s prekomernom težinom u ko		
1897	Ove crte njegove ličnosti u velikoj su meri doprinele njegovoj slavi nakon smrti.		
1898	Zahvaljujući žrtvovanju	Isusa	na krstu, ljudski gresi su oprošteni, ali njihova grešna priroda n
1899	Isto važi i za	F-35B Lightning II	, koji je pokazao sposobnost VTOL u probnim letovima, ali je u o
1900	Parfemi su čest uzrok akutnih napada kod žena i dece.		
1901	Po konvenciji, standardni oblik definiše problem minimizacije.		
1902	Elektrokardiogram se koristi da bi se utvrdilo da li je srce pod opterećenjem zbog	visokog krvnog pritiska.	

Named entity
"Evropi"
ID:725088
identifier: <http://www.wikidata.org/entity/Q46>
Label: Europe
Description: continent

Gr.	Type	N _f	N _l
MWE	NOUN	413	303
	PNOUN	136	73
	ADV	45	21
	ADJ	2	2
VMWE	IRV	205	83
	LVC _{FULL}	43	28
	VID	12	9
	LVC _{CAUSE}	10	7

A data model for a parallel corpus in NIF



Stanković, Ranka, Ikonić Nešić
Milica, Olja Perišić, Mihailo
Škorić, and Olivera Kitanović.

"Towards Semantic Interoperability: Parallel Corpora as Linked Data Incorporating Named Entity Linking."

In The 9th Workshop on Linked
Data in Linguistics: Resources,
Applications, Best Practices
(LDL-2024)@
LREC-COLING-2024, pp.
115-125. [pdf](#)

Lexical Entry #216633

krvni pritisak

Check in dictionaries:

- [show RSinonima](#)
- [show Terminološki](#)

Check in external dictionaries: [Wiktionary](#) [Babelnet](#) [Termi](#) [Glosbi](#)Search corpora: [Concordances](#) [Attestations](#) [Form Frequen](#)

Senses (1):

1. +Comp+DOM=Med

Domains:	medicina
Properties:	polileksemska jedinica
Note:	Danas;mar07

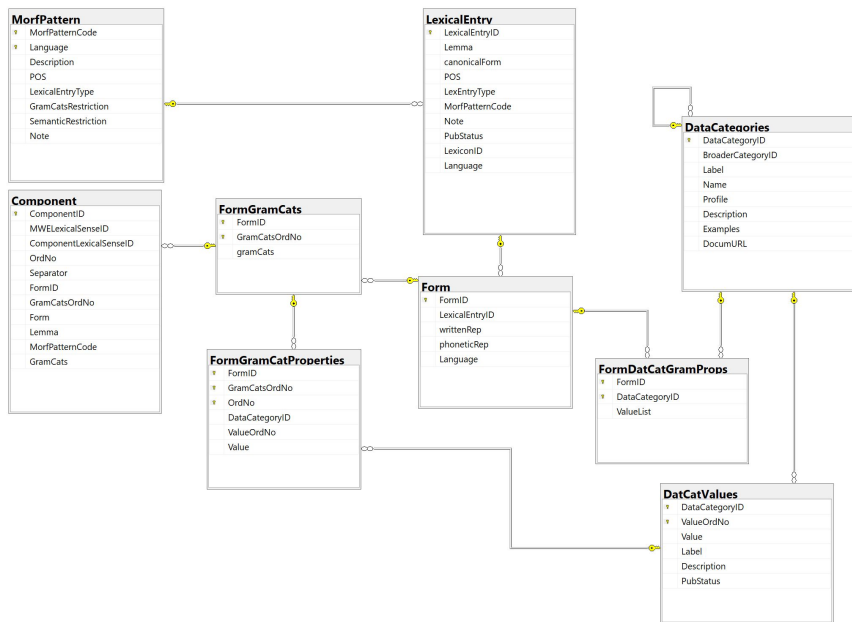
Is composed of:

Form	Lemma	FST Code	Gram Cat	Separator
krvni	krvni	A2	adms1g	
pritisak	pritisak	N17	ms1q	

krvna pritiska,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:mw2q:mw4q
krvne pritiske,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:mp4q
krvni pritisak,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms1q:ms4q
krvni pritisaci,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:mp1q:mp5q
krvni pritisče,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms5q
krvnih pritisaka,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:mp2q
krvnim pritiscima,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:mp3q:mp6q:mp7q
krvnim pritiskom,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms6q
krvnima pritiscima,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:mp3q:mp6q:mp7q
krvnog pritiska,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms2q
krvnoga pritiska,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms2q
krvnom pritisku,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms3q:ms7q
krvnome pritisku,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms3q:ms7q
krvnomu pritisku,krvni(krvni.A2:adms1g) pritisak(pritisak.N17:ms1q).NC:ms3q:ms7q

Ontolex-lemon

Agata Savary, Cvetana Krstev, Duško Vitas, “Inflectional non compositionality and variation of compounds in French, Polish and Serbian, and their automatic processing”, in Bulag - Bulletin de Linguistique Appliquée et Générale, Les langues slaves et le français : approches formelles dans les études contrastives, No. 32, pp. 73-94, Presses Universitaires de Franche Comté, Besancon, 2007.



Ranka Stanković, Cvetana Krstev, Biljana Lazić and Mihailo Škorić, “Electronic Dictionaries - from File System to lemon Based Lexical Database”, LREC2018 - W23 6th Workshop on Linked Data in Linguistics : Towards Linguistic Data Science (LDL-2018), Japan, 2018.

```

ontolex:entry :le_pritisak_16192, :le_krvni_78511,
               :le_krvni_pritisak_216633.
:cm_krvni_11318 decomp:correspondsTo :le_krvni_78511.
:cm_pritisak_11319 decomp:correspondsTo :le_pritisak_16192.
:fm_krvna_5289787 a ontolex:Form; ontolex:writtenRep "крвна"@sr.
:fm_krvna_pritiska_2265797 a ontolex:Form; ontolex:writtenRep "крвна притиска"@sr.
:fm_krvne_5289788 a ontolex:Form; ontolex:writtenRep "крвне"@sr.
...
:fm_pritisak_4432590 a ontolex:Form; ontolex:writtenRep "притисак"@sr.
:fm_pritisaka_4432591 a ontolex:Form; ontolex:writtenRep "притисака"@sr.
:fm_pritisaci_4432592 a ontolex:Form; ontolex:writtenRep "притисци"@sr.
...
:fm_krvne_pritiske_2265798 a ontolex:Form; ontolex:writtenRep "крвне притиске"@sr.
fm_krvni_pritisak_2265799 a ontolex:Form; ontolex:writtenRep "крвни притисак"@sr.
:fm_krvni_pritisaci_2265800 a ontolex:Form; ontolex:writtenRep "крвни притисци"@sr.
...
:le_krvni_78511 olia:partOfSpeech olia:Adjective;
lexinfo:partOfSpeech lexinfo:adjective;
a ontolex:LexicalEntry, ontolex:Word;
ontolex:canonicalForm [ontolex:writtenRep "крвни"@sr];
ontolex:lexicalForm :fm_krvna_5289787, :fm_krvne_5289788, :fm_krvni_5289789,...

:le_pritisak_16192 olia:partOfSpeech olia:CommonNoun;
lexinfo:partOfSpeech lexinfo:commonNoun;
a ontolex:LexicalEntry,
  ontolex:Word;
ontolex:canonicalForm [ontolex:writtenRep "притисак"@sr];
ontolex:lexicalForm :fm_pritisak_4432590, :fm_pritisaka_4432591,...

:le_krvni_pritisak_216633 olia:partOfSpeech <http://purl.org/olia/olia.owl/#->;
lexinfo:partOfSpeech lexinfo:ProperNoun;
rdf:_1 :le_krvni_78511;
rdf:_2 :le_pritisak_16192;
a ontolex:LexicalEntry, ontolex:MultiwordExpression, decomp:Component;
decomp:constituent :cm_krvni_11318, :cm_pritisak_11319;
ontolex:canonicalForm [ontolex:writtenRep "крвни притисак"@sr];
ontolex:lexicalForm :fm_krvna_pritiska_2265797, :fm_krvne_pritiske_2265798,...
  
```

```

:le_fast_food
  a ontolex:LexicalEntry, ontolex:MultiwordExpression;
  ontolex:canonicalForm [ontolex:writtenRep "fast food"@en];
  lexinfo:partOfSpeech lexinfo:noun;
  ontolex:sense
    [ontolex:reference <https://www.wikidata.org/wiki/Q81799>];
  decomp:constituent :cm_fast;
  decomp:constituent :cm_food;
  rdf:_1 :le_fast; # lexical
  rdf:_2 :le_food. # entries

```

component of canonical form

```

:cm_food a decomp:Component;
  decomp:correspondsTo :le_food.

```

...

```

:le_brza_hrana a ontolex:LexicalEntry, ontolex:MultiwordExpression;
  ontolex:canonicalForm [ontolex:writtenRep "brza hrana"@sr];

```

...

simplified naming

```

:tranSetEN-SR vartrans:trans fast_food-ensns-brza_hrana-srsns .
:fast_food-ensns a ontolex:LexicalSense ;
  ontolex:isSenseOf :le_fast_food .
:brza_hrana-srsns a ontolex:LexicalSense ;
  ontolex:isSenseOf :le_brza_hrana .
:fast_food-ensns-brza_hrana-sns-trans a vartrans:Translation ;
  vartrans:source :fast_food-ensns ;
  vartrans:target :brza_hrana-srsns .

```

```

:le_fast_food
  frac:attestation [
    frac:quotation "It can be made at home or
bought from fast fooshops."@en;
    frac:observedIn :EWSD].

```

```

:le_brza_hrana
  frac:attestation [
    frac:quotation "Može se napraviti
kod kuće ili kupiti u prodavnicama
brze hrane."@sr;
    frac:observedIn :EWSD-ext].

```

Cvetana Krstev, Ranka Stanković, Aleksandra M. Marković, and Teodora Sofija Mihajlov. 2024. **Towards the semantic annotation of SR-ELEXIS corpus: Insights into Multiword Expressions and Named Entities**. Proc. of the Joint Workshop on Multiword Expressions and Universal Dependencies (MWE-UD) @ LREC-COLING 2024, Torino.

Wikidata - VMWE classes

LVC.cause Q131916794

LVC.full Q131916737

VID Q131916805

~ ^ 131916809

```
<https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=1592,1614> a nif:Phrase,  
    nif:RFC5147String,  
    nif:String ;  
nif:anchorOf "Pristup imaju"^^xsd:string ;  
nif:beginIndex "1592"^^xsd:nonNegativeInteger ;  
nif:endIndex "1614"^^xsd:nonNegativeInteger ;  
nif:referenceContext <https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt> ;  
nif:sentence <https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=1592,1685> ;  
dc:Language "sr"^^xsd:string ;  
itsrdf:taClassRef wd:Q131916737 .
```

```
<https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=159320,159329> a nif:Phrase,  
    nif:RFC5147String,  
    nif:String ;  
nif:anchorOf "se dogodi"^^xsd:string ;  
nif:beginIndex "159320"^^xsd:nonNegativeInteger ;  
nif:endIndex "159329"^^xsd:nonNegativeInteger ;  
nif:referenceContext <https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt> ;  
nif:sentence <https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=159249,159361> ;  
dc:Language "sr"^^xsd:string ;  
itsrdf:taClassRef wd:Q131916809 .
```

```

1 PREFIX itsrdf: <http://www.w3.org/2005/11/its/rdf#>
2 PREFIX wd: <http://www.wikidata.org/entity/>
3 PREFIX nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
4
5 SELECT ?subject ?anchor WHERE {
6   ?subject itsrdf:taClassRef wd:Q131916737 ;
7     nif:anchorOf ?anchor .
8 }

```

Table Response 49 results in 0.036 seconds

	subject	anchor
1	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=100650,100662>	Ime dobio
2	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=11394,11404>	dobila ime
3	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=116091,116108>	imaju šanse
4	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=118755,118764>	dobio ime
5	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=124707,124721>	imaju problema
6	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=124921,124939>	izvrši samoubistvo
7	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=13052,13073>	dao doprinos
8	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=13313,13330>	pruža informacije
9	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=133256,133265>	ima pravo
10	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=137428,137453>	imaju podršku
11	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=14380,14390>	sastaju se
12	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=14449,14464>	vodili raspravu
13	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=14654,14670>	iznese stav

<http://fuseki.jerteh.rs/#/dataset/sr-ELEXIS-U/NIDIVE-demo/query>

PREFIX itsrdf:

<http://www.w3.org/2005/11/its/rdf#>

PREFIX wd: <http://www.wikidata.org/entity/>

PREFIX nif:

<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>

SELECT ?subject ?anchor WHERE {

?subject itsrdf:taClassRef wd:Q131916737 ;

nif:anchorOf ?anchor .

}

<http://fuseki.jerteh.rs/#/dataset/sr-ELEXIS-UNIDIVE-demo/query>

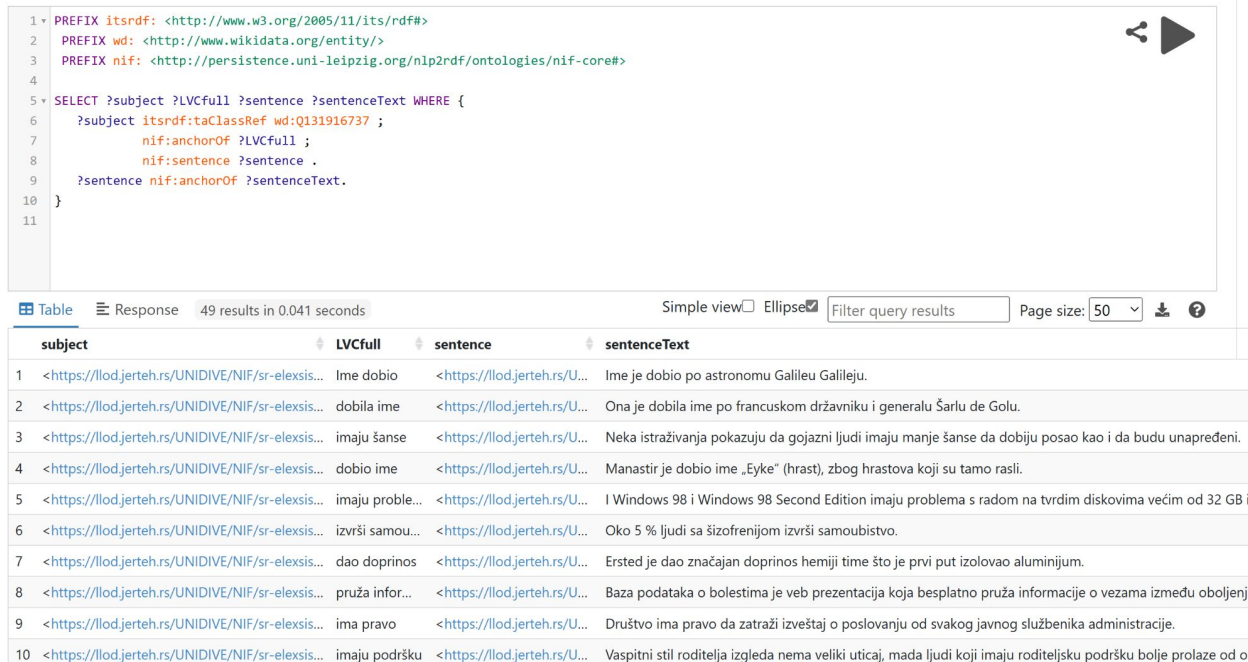
PREFIX wd:

<http://www.wikidata.org/entity/>

PREFIX nif:

<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>

```
SELECT ?subject ?LVCfull ?sentence
?sentenceText WHERE {
  ?subject itsrdf:taClassRef
wd:Q131916737 ;
  nif:anchorOf ?LVCfull ;
  nif:sentence ?sentence .
?sentence nif:anchorOf ?sentenceText.
}
```



The screenshot shows the Fuseki query interface. At the top, there is a text area containing a SPARQL query. Below the text area, there are tabs for 'Table', 'Response', and '49 results in 0.041 seconds'. The 'Table' tab is selected, displaying a table with 4 columns: 'subject', 'LVCfull', 'sentence', and 'sentenceText'. The table contains 10 rows of results, each with a row number and truncated values for the first three columns. The 'sentenceText' column contains the full text of the sentences.

	subject	LVCfull	sentence	sentenceText
1	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	Ime dobio	<https://lloj.jerteh.rs/U...	Ime je dobio po astronomu Galileu Galileju.
2	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	dobila ime	<https://lloj.jerteh.rs/U...	Ona je dobila ime po francuskom državniku i generalu Šarlu de Golu.
3	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	imaju šanse	<https://lloj.jerteh.rs/U...	Neka istraživanja pokazuju da gojazni ljudi imaju manje šanse da dobiju posao kao i da budu unapređeni.
4	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	dobio ime	<https://lloj.jerteh.rs/U...	Manastir je dobio ime „Ejke“ (hrast), zbog hrastova koji su tamo rasi.
5	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	imaju probleme	<https://lloj.jerteh.rs/U...	I Windows 98 i Windows 98 Second Edition imaju problema s radom na tvrdim diskovima većim od 32 GB i
6	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	izvrši samou...	<https://lloj.jerteh.rs/U...	Oko 5 % ljudi sa šizofrenijom izvrši samoubistvo.
7	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	dao doprinos	<https://lloj.jerteh.rs/U...	Ersted je dao značajan doprinos hemiji time što je prvi put izolovao aluminijum.
8	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	pruža infor...	<https://lloj.jerteh.rs/U...	Baza podataka o bolestima je veb prezentacija koja besplatno pruža informacije o vezama između oboljenja
9	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	ima pravo	<https://lloj.jerteh.rs/U...	Društvo ima pravo da zatraži izveštaj o poslovanju od svakog javnog službenika administracije.
10	<https://lloj.jerteh.rs/UNIDIVE/NIF/sr-elexis...	imaju podršku	<https://lloj.jerteh.rs/U...	Vaspitni stil roditelja izgleda nema veliki uticaj, mada ljudi koji imaju roditeljsku podršku bolje prolaze od ro

It-Sr-NER - aligned sentences

PREFIX wd:

<http://www.wikidata.org/entity/>

PREFIX nif:

<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>

PREFIX itsrdf:

<http://www.w3.org/2005/11/its/rdf#>

PREFIX skos:

<http://www.w3.org/2004/02/skos/core#>

SELECT ?sr ?srt ?it ?itt

WHERE { ?sr a nif:Sentence;

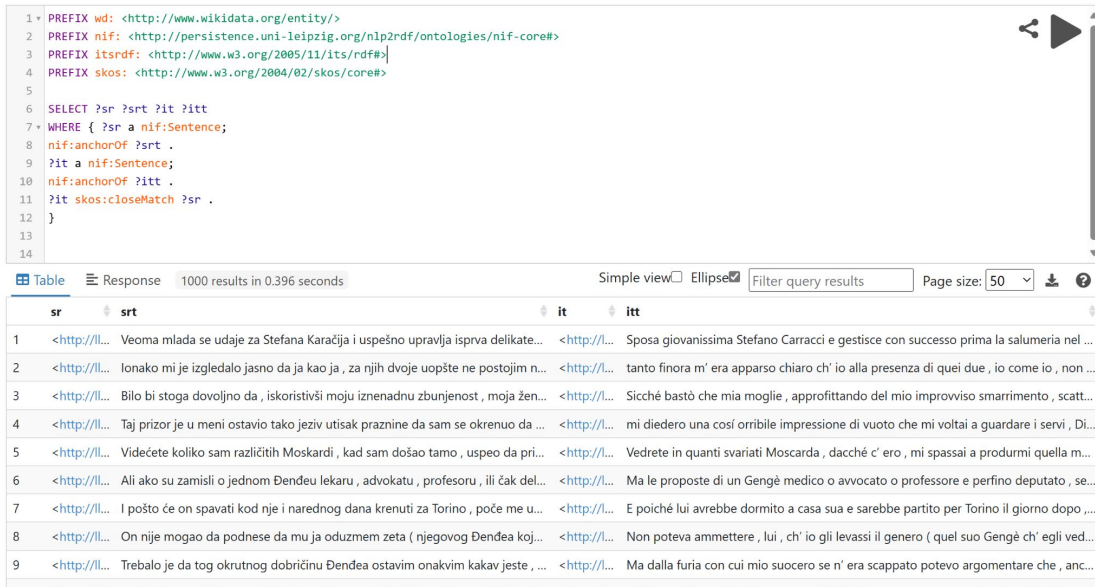
nif:anchorOf ?srt .

?it a nif:Sentence;

nif:anchorOf ?itt .

?it skos:closeMatch ?sr .

}



The screenshot shows a SPARQL query interface. The query is as follows:

```
1 PREFIX wd: <http://www.wikidata.org/entity/>
2 PREFIX nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
3 PREFIX itsrdf: <http://www.w3.org/2005/11/its/rdf#>
4 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5
6 SELECT ?sr ?srt ?it ?itt
7 WHERE { ?sr a nif:Sentence;
8         nif:anchorOf ?srt .
9         ?it a nif:Sentence;
10        nif:anchorOf ?itt .
11        ?it skos:closeMatch ?sr .
12      }
13
14
```

The results are displayed in a table with 4 columns: sr, srt, it, and itt. The table shows 9 rows of results, each representing a pair of aligned sentences in Serbian (sr) and Italian (itt).

	sr	srt	it	itt
1	<http://l...	Veoma mlada se udaje za Stefana Karačija i uspešno upravlja isprva delikate...	<http://l...	Sposa giovanissima Stefano Carracci e gestisce con successo prima la salumeria nel ...
2	<http://l...	Ionako mi je izgledalo jasno da ja kao ja , za njih dvoje uopšte ne postojim n...	<http://l...	tanto finora m' era apparso chiaro ch' io alla presenza di quei due , io come io , non ...
3	<http://l...	Bilo bi stoga dovoljno da , iskoristivši moju iznenadnu zbunjenost , moja žen...	<http://l...	Sicché bastò che mia moglie , approfittando del mio improvviso smarrimento , scatt...
4	<http://l...	Taj prizor je u meni ostavio tako jeziv utisak praznine da sam se okrenuo da ...	<http://l...	mi diedero una così orribile impressione di vuoto che mi voltai a guardare i servi , Di...
5	<http://l...	Videćete koliko sam različitih Moskardi , kad sam došao tamo , uspeo da pri...	<http://l...	Vedrete in quanti svariati Moscarda , dacché c' ero , mi spassai a produrmi quella m...
6	<http://l...	Ali ako su zamisli o jednom Dendeu lekaru , advokatu , profesoru , ili čak del...	<http://l...	Ma le proposte di un Gengè medico o avvocato o professore e perfino deputato , se...
7	<http://l...	I pošto će on spavati kod nje i narednog dana krenuti za Torino , poče me u...	<http://l...	E poiché lui avrebbe dormito a casa sua e sarebbe partito per Torino il giorno dopo ,...
8	<http://l...	On nije mogao da podnese da mu ja oduzmem zeta (njegovog Dendea koj...	<http://l...	Non poteva ammettere , lui , ch' io gli levassi il genero (quel suo Gengè ch' egli ved...
9	<http://l...	Trebalo je da tog okrutnog dobičinu Dendea ostavim onakvim kakav jeste , ...	<http://l...	Ma dalla furia con cui mio suocero se n' era scappato potevo argomentare che , anc...

Uncontinuous MWEs ?

23 Mladi journalist | ROLE novinari 1 | pres-centar | NOUN pres centra 2 | iznositi | LVCfull iznose 2 | viđenje | LVCfull svoje viđenje učešća u Programu, realizujući format „deca o deci“.

```
<https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=2288,2308> a nif:Phrase,  
    nif:RFC5147String,  
    nif:String ;  
nif:anchorOf "iznose viđenje"^^xsd:string ;  
nif:beginIndex "2288"^^xsd:nonNegativeInteger ;  
nif:endIndex "2308"^^xsd:nonNegativeInteger ;  
nif:referenceContext <https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt> ;  
nif:sentence <https://llo.d.jerteh.rs/UNIDIVE/NIF/sr-elexsis.txt#char=2261,2361> ;  
dc:Language "sr"^^xsd:string ;  
itsrdf:taClassRef wd:Q131916737 .
```

sr-ELEXIS sentences+
NER+NEL+
geolocations
as
geoJSON

```

  [0] {3}
    type : "Feature"
    properties {8}
      QID : "Q3766"
      name : "Damascus"
      description : "capital and largest city of Syria"
      lat : 33.513055555556001
      lon : 36.291944444443999
      type : "city, big city, populated place in Syria, national capital, largest city"
      entity_text : "Damasku"
      full_sentence : "Više od 7000 ljudi otišlo je na premijeru filma u Damasku ."
    geometry {2}
      type : "Point"
      coordinates [2]
        [0] : 36.291944444443999
        [1] : 33.513055555556001
  [1] {3}

```

sr-ELEXIS
NER+NEL+
sentences+
gelocations
as
geoJSON
to
QGIS

QGIS interface showing a map of Europe with various locations marked. The left sidebar displays the 'Browser' panel with a file tree and the 'Layers' panel with a list of layers. The 'Locations (14)' layer is selected, showing 621 features. The 'Expression' panel is open, displaying a table of data for the selected layer.

Browser

- Busorinetest.gpkg
 - busotine_test.shp
 - buzet.pdf
 - D2.1 Report on the tra
 - D4.1 Social networkin
 - D4.2 Workshop and cc
 - locations (14).geojson
 - locations.geojson
 - ProgramJudig.pdf
 - ProgramOverviewJudig
 - put.shp
 - query (1).json

Layers

- locations (14) [621]
- OpenStreetMap

locations (14) — Features Total: 621, Filtered: 621, Selected: 0

Expression

☐ Italy	☐ North America
☐ Provence	☐ Germany
☐ Winniepeg	☐ Europe
☐ Denmark	☒ Venice
☐ Sarajevo	☐ Atrebatas
☐ Britain	☐ United States

QID Q641
name Venice
description capital city of Veneto, Italy
lat 45,439722222222
lon 12,331944444444
type commune of Italy, carfree city, port settlement, city, big city, Italian city-state, car-free place
entity_text Venecija
full_sentence Nakon te bitke borbe su se ponovo prenele na francusko tlo , dok je Venecija sklopila separadni mir

Challenges and Next steps

- Verbosity of RDF
 - txt (226 KB), tsv (1.9 MB), ttl (27 MB), fuseki (243 MB)
- Scalability
 - for large corpora, indexes grow fast
- Uncontinuous MWEs
- Good endpoint
 - better RDF store - Virtuoso?
- DMLex
- MWE Ontology