

Leveraging Linked Data, NIF, and CONLL-U for Enhanced Annotation in Sentence Aligned Parallel Corpora



WG2

Otolex-lemon
Decomp
Vratrans
FRaC

- Interlinking MWE lexicon entries with their occurrences in corpora
- publishing of aligned and annotated corpus as LD employing NIF

ELEXIS-WSD
Parallel
Sense-
Annotated
Corpus

```
<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>.
```

```
]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
```