

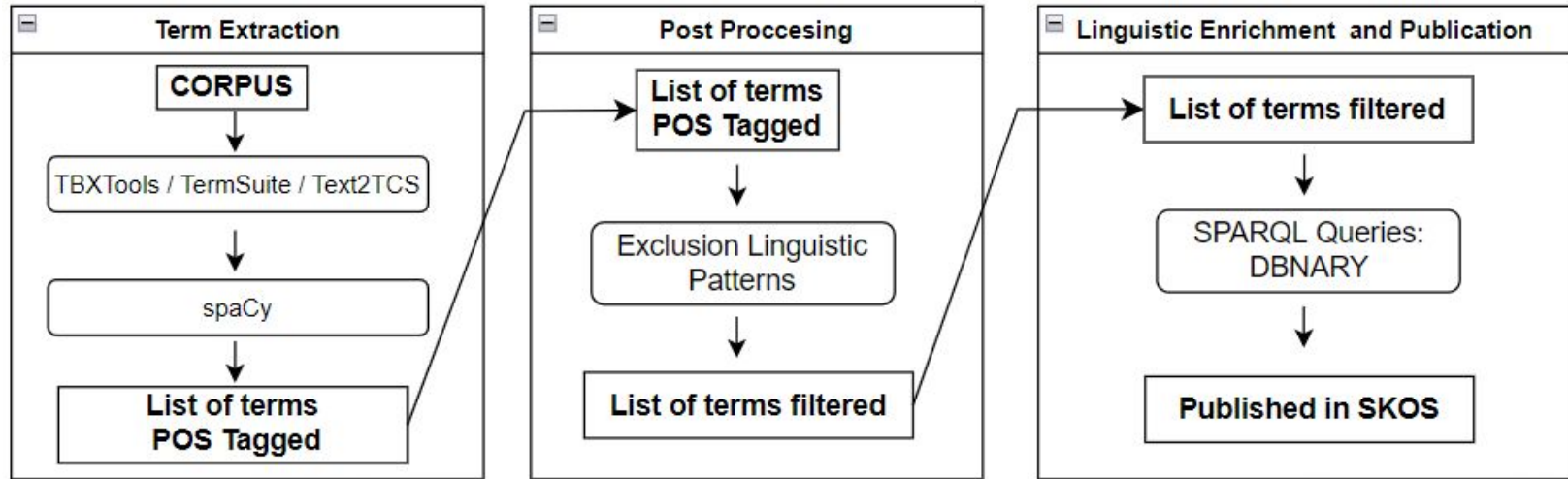


Biomedical Terminology

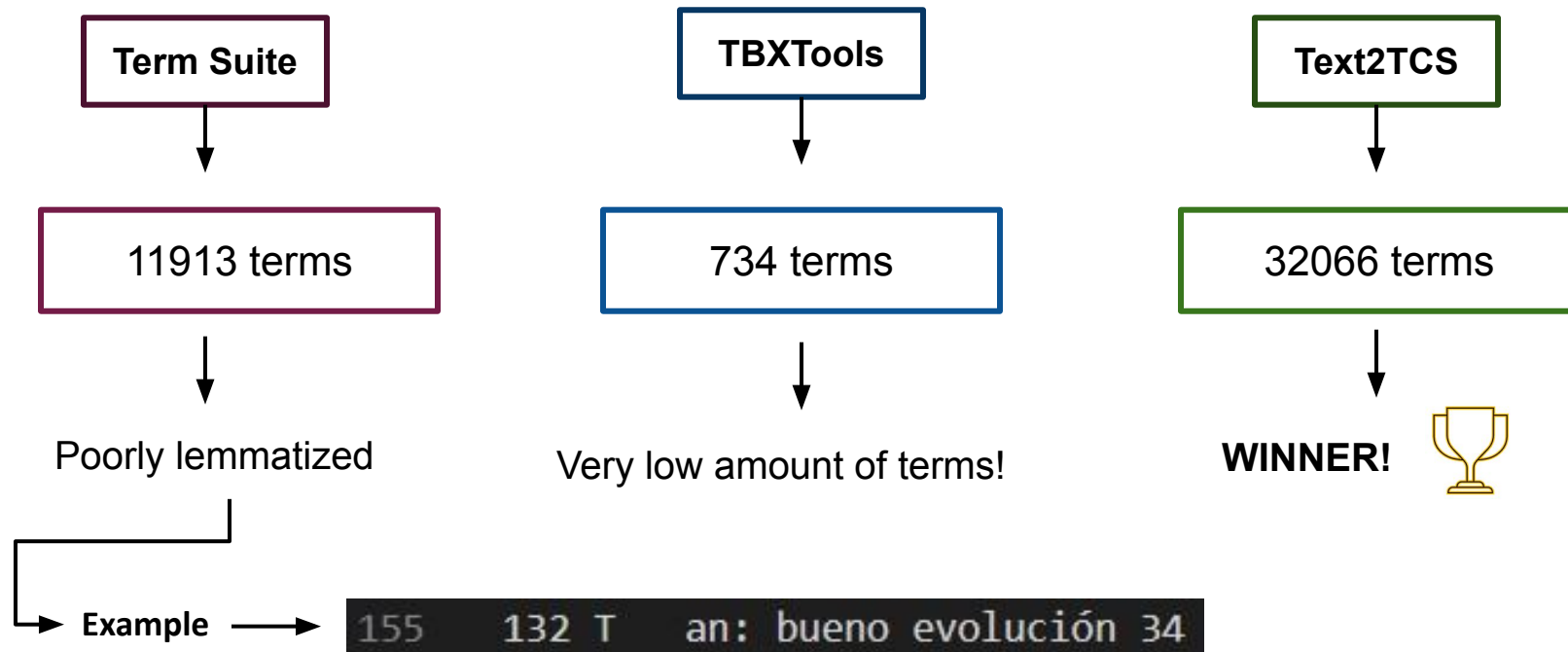
Lucia Sanchez, Carlos Golvano and Renzo Alva Principe
Tutor: Dagmar Gromann



Overview



Terminology Extraction



Post-processing

- 1º Remove [most frequent general Spanish words](#) → **633 of 32066 terms deleted**
- 2º **spaCy** → POS Tags → Remove words using linguistic patterns. Final result: 8687 terms

Exclusion patterns	Example
adjective + noun + adjective	renal tipo papilar
adverb	previo
adjective	alto
verb + adjective	mandibular derecha

Linguistic Enrichment



MeSH
SNOMED CT
Other Bioportal sources
Datos.gob



No data in Spanish



DBnary



Translations

Linked Data Modelling

- ❖ **SKOS** based
- ❖ relations:
 - SKOS:inScheme
 - SKOS:closeMatch
 - SKOS:prefLabel

```
<http://sd-llod-22.org/medterm/es_viruela> a SKOS:Concept .
<http://sd-llod-22.org/medterm/es_viruela> SKOS:inScheme "disease" .
<http://sd-llod-22.org/medterm/es_viruela> SKOS:closeMatch dbnary-spa:viruela__sustantivo_femenino__1 .
<http://sd-llod-22.org/medterm/es_viruela> SKOS:prefLabel "brec'h@br" .
<http://sd-llod-22.org/medterm/es_viruela> SKOS:prefLabel "smallpox@en" .
<http://sd-llod-22.org/medterm/es_viruela> SKOS:prefLabel "variole@fr" .
<http://sd-llod-22.org/medterm/es_viruela> SKOS:prefLabel "ospa@pl" .
```

Project Results

- NLP pipeline for bio-medical terminology extraction
- https://github.com/rAlvaPrincipe/med_terminology

Med
TermitUp



Future Work

- Enrich with more linguistic linked data of the biomedical domain
- Represent additional data with OntoLex model
- Integration of the pipeline in TermitUp