



4th Summer Datathon on Linguistic Linked Open Data

Cercedilla, Spain

30th May – 3 June 2022



# MinuteMadness



## INSTRUCTIONS FOR THE MINUTE MADNESS:

- This is an opportunity for participants to introduce themselves and share their backgrounds, motivations and maybe pointers to relevant work with everyone else
- I will read out each participant's name and affiliation and show their pre-prepared slide. The participant will then have **one minute** to talk about themselves

Later, during this session, the main proposer of every “miniproject” will have up to five minutes to briefly motivate and present their project proposal.

Without any further ado I present...

**Lucía Pitarch  
Ballesteros**

***Universidad de Zaragoza***

## Researcher and Writer, Computational Linguist



## Main research line and interests

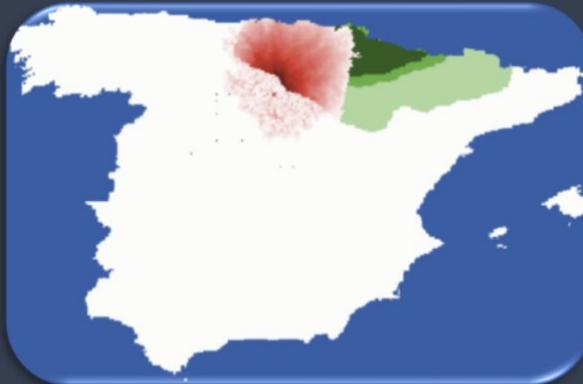
- Thesis on Multilingual Linguistic Linked Data
- Currently focusing on relation extraction
- Interests: Multilingualism, Language contact, Language change and their computational study

## Some current and previous work

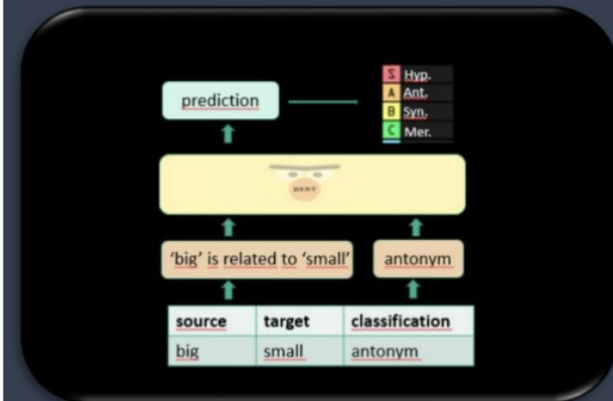
How senses are grouped within word forms over time: an analysis of diachronic colexification patterns.



Modelling phonetic change. A case study on Latin's initial F- loss

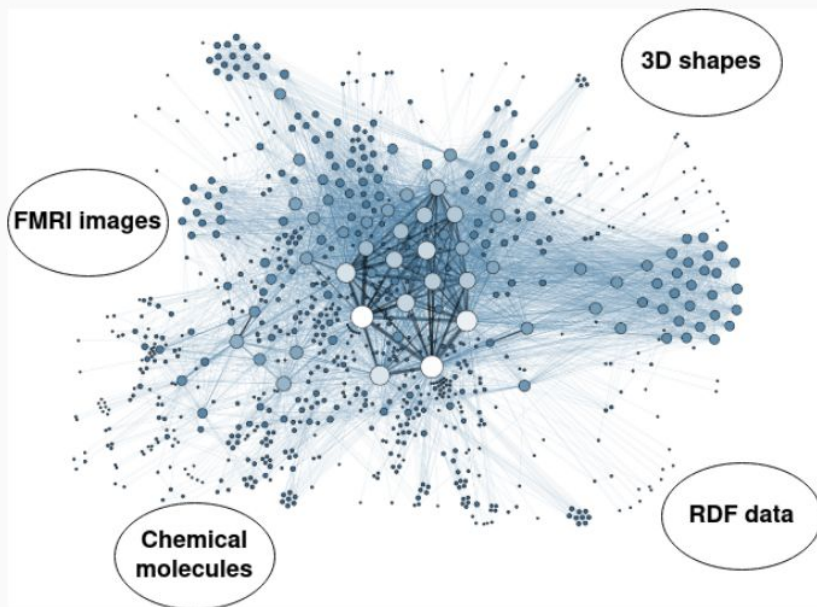


Lexico-semantic relation classification with multilingual finetuning



**Katerina  
Gkirtzou**

***Athena Research  
Center/Institute for  
Language and Speech  
Processing***



openMIN7ED  
Open Mining Infrastructure for Text & Data

clarin:el

EUROPEAN  
LANGUAGE  
GRID

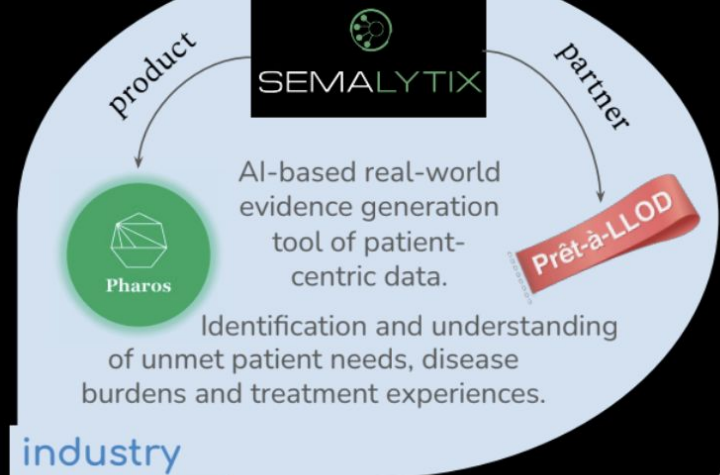
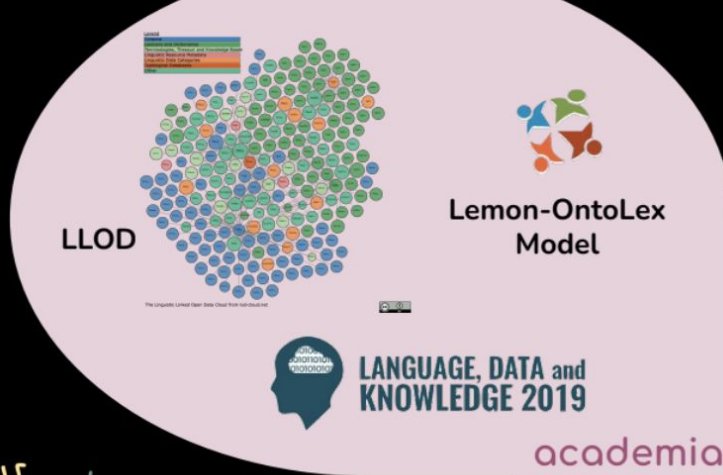
**Danielle Gregg**  
***Universität des***  
***Saarlandes***

# Danielle Gregg

- BS in Mathematics & Computer Science
  - Union College, Schenectady, New York, USA
- MSc in Language Science & Technology
  - Universität des Saarlandes, Saarland, Germany
- Interested in:
  - Translation & translation technologies
  - Language preservation
  - Equitable access to technology
- Besides academia:
  - Dancing
  - Gaming - Skyrim & Animal Crossing
  - Rupaul's Drag Race & Survivor



**Bettina Klimek**  
***Semalytix***



JOIN THE PROJECT



Mini project: **SD-LLOD-2022**  
"Creating a wordnet of medical concepts"  
(including scientific discussions about 'hangover', 'forgetting' and 'bizarre behavior')

INTENSE BUT FUN



#noentityleftbehind



SEMALYTIX

**Ibai Guillén**

***Universidad  
Politécnica de  
Madrid***



Universidad de Deusto  
Deustuko Unibertsitatea  
University of Deusto

**Bachelor's degree**

Computer Engineering +  
Business Digital  
Transformation



**POLITÉCNICA**

UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

**Master of Science**

ARTIFICIAL  
INTELLIGENCE



Ibai Guillén Pacho



**Research intern**

Tracking open datasets through  
Blockchain



**Internship**

Topic modelling



**Santurtzi**

Biscay, Basque Country

**Carlos Golvano  
Díaz**

***Polytechnic  
University of Madrid***



Carlos Golvano Díaz



Universidad  
de Alcalá

Computer Engineering



POLITÉCNICA

UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

Artificial Intelligence



Guadalajara, Spain



Master's thesis  
Natural Language Processing

**Giedre Valunaite**  
**Oleskeviciene**  
***Mykolas Romeris***  
***University***



# Giedrė Valūnaitė Oleškevičienė MYKOLAS ROMERIS UNIVERSITY LITHUANIA



**Prof. Dr. Giedrė Valūnaitė Oleškevičienė** Professor at the Institute of Humanities

Works in the domains of humanities and social sciences. In the domain of humanities, her scientific interests include discourse analysis, discourse annotated corpora, professional English and legal English, linguistics and translation research. The researcher is also actively engaged in second language teaching and learning research.

Publications are available:  
[https://www.researchgate.net/profile/Giedre\\_Valunaite\\_Oleskeviciene/research](https://www.researchgate.net/profile/Giedre_Valunaite_Oleskeviciene/research)



**In NexusLinguarum** works on UC4.2.2 Social Sciences exploring discourse markers identifying speaker attitude, also UC4.1.1 Media and Social Media focusing on offensive language identification and many other groups.



**Datathon expectations** – to continue working on BATS initiative And also a **MAD expectation** – we have the data collected for Lithuanian Rhetorical figures and are looking for opportunities to link the LT data to the existing resources of rhetorical figures in other languages





**Renzo Arturo**  
**Alva Principe**  
*University of*  
*Milano-Bicocca*

# Hi! I'm Renzo Alva Principe

Ph.D. student in NLP @ *University of Milano-Bicocca* (1st year)

- before my Ph.D.
  - Scalable Knowledge Graph Profiling
  - Neural Relation Extraction
- during my Ph.D.
  - Classification of Digital Documents
  - Information Extraction (rule-based at the moment)
- motivation and dream
  - structure the whole human knowledge contained in text sources in a scalable way
  - get in touch with people with similar aspirations



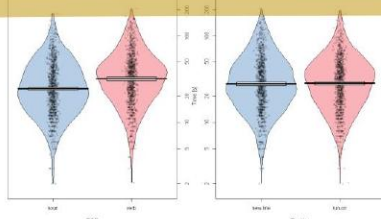
**Robert Lew**  
***Adam Mickiewicz***  
***University***



rakieta

f.

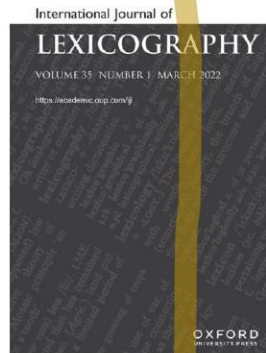
1. (*pojazd*) rocket; **rakieta kosmiczna** (space) rocket; **rakieta nośna** launch vehicle
2. **wojsk** (*pośisk*) rocket, missile
3. (= *raca*) flare, signal rocket, Very lights.
4. **sport** r...
5. (= *sprz*) snowsł



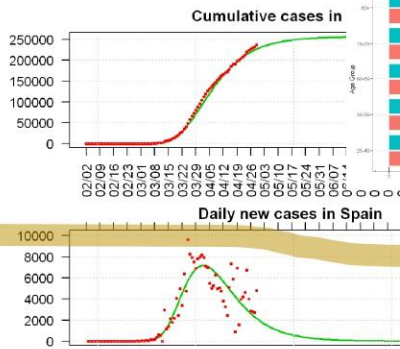
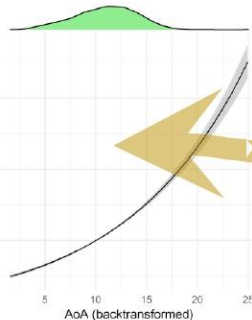
EUROPEAN NETWORK  
OF e-LEXICOGRAPHY



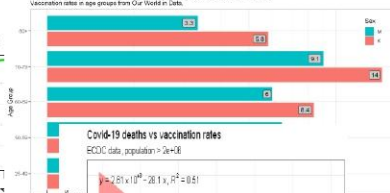
european lexicographic  
infrastructure



Predicted views (backtransformed)

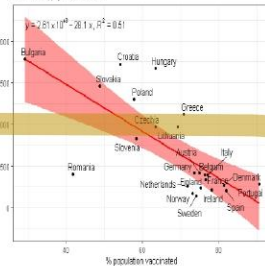


Covid death risk ratio, fully vaccinated vs all else, Poland  
Deaths with Covid between 2019-05-01 and 2022-03-14  
<https://covid19.gis.polimex.gov.pl/eng/analiza-wiekow-i-powodu-zmarli-19-a-wiek>  
Vaccination rates in age groups from Our World in Data.



Covid-19 deaths vs vaccination rates

ECDC data, population > 2m



colloCaid

**Kristina Kocijan**

***Faculty of  
Humanities and  
Social Sciences,  
University of Zagreb***



University of Zagreb

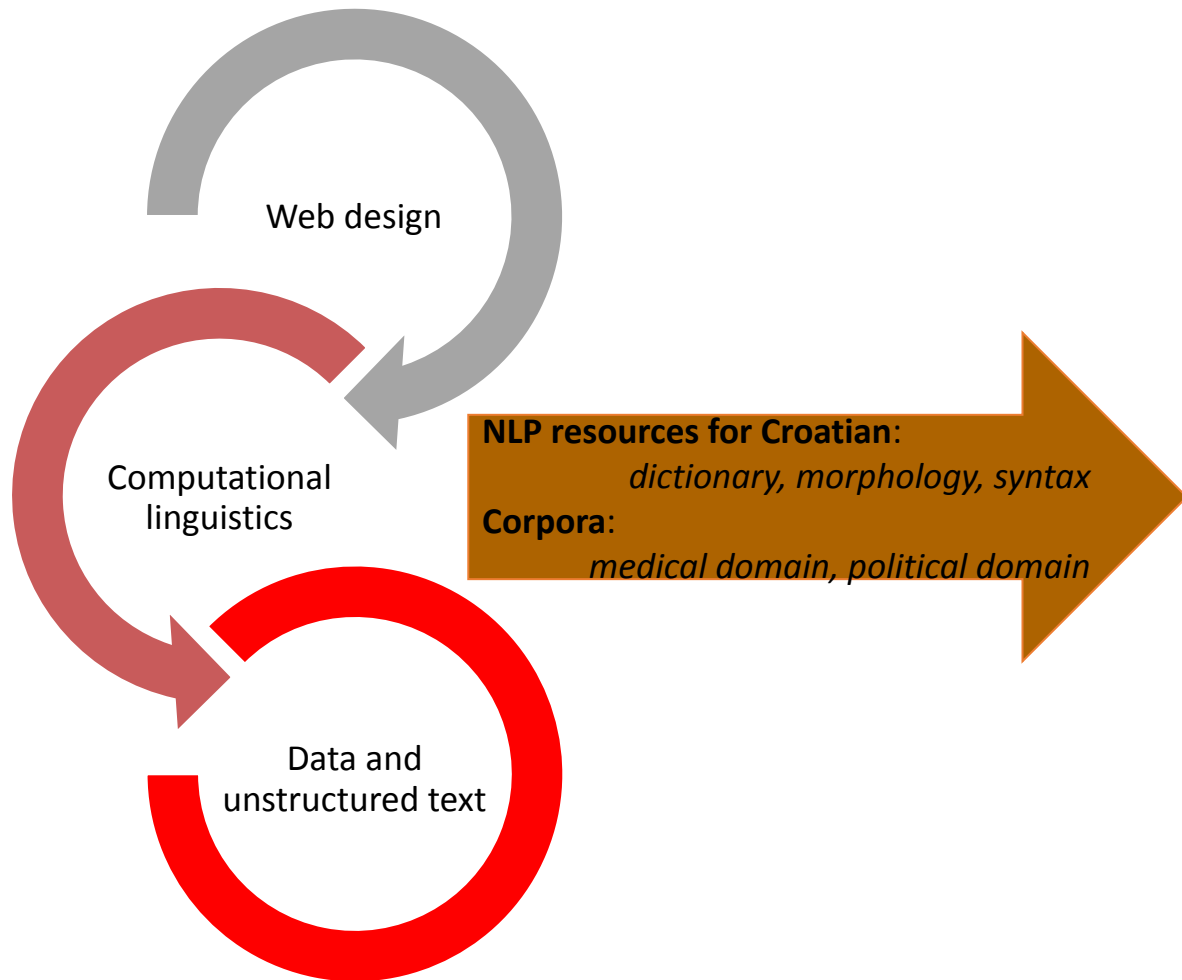


Faculty of Humanities and Social  
Sciences



Department of Information and  
Communication Sciences

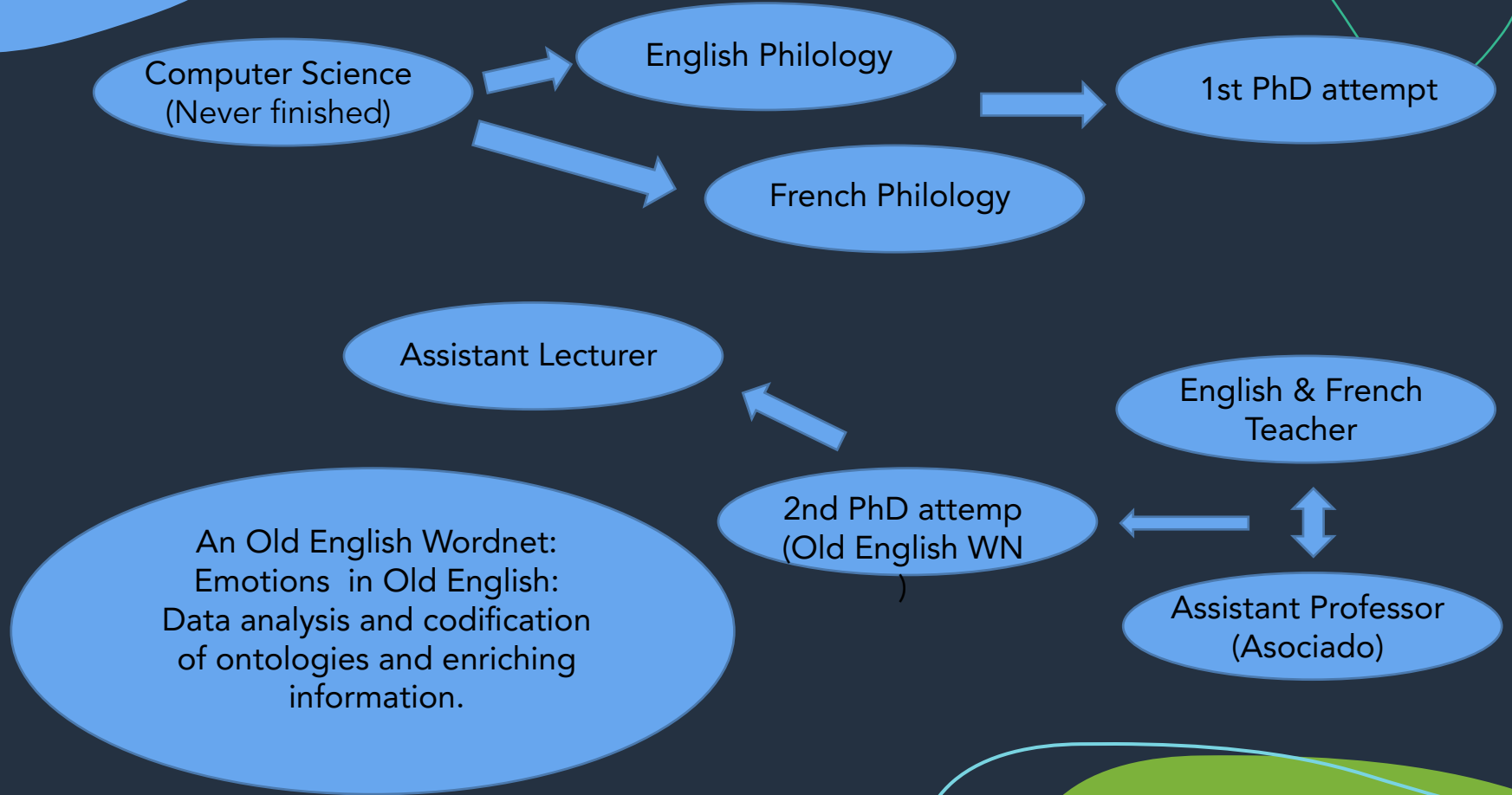
*Kristina Kocijan,*  
*Associate Professor*



**Rafael Cruz**

**González**

***UCLM***





**Khadīja Ait**

**ElFqih**

***University of Naples***

***L'Orientale***

## **Academic Background**

- PhD student in the department of literary, Linguistic and Comparative Studies and member of the UNIOR NLP Research Group at the university of Naples l'Orientale. My research evolves into the Complexities of translating Legal Terms from Arabic into English, Terminology Management, and Standardization. I am working on building a multilingual dictionary (Arabic-English/French-Italian) of legal terms available in different standard and digital formats.
- Master's degree of Translation Technology and specialised translation at Cadi Ayyad University, Faculty of letters and human sciences in Marrakech, Morocco.
- Erasmus+ Exchange Researcher in literary, Linguistic and Comparative Studies department at the university of Naples L'Orientale.
- I have completed a co-tutored master thesis in 'The Impact of terminological Complexities in Translating Legal Terms from Arabic into English : Marriage contract and Divorces Provision as a case of Study' In Cadi Ayyad University the the university of Naples l'Orientale. Besides to many other projects and internships related to terminology and translation within these two years program.
- General academic Studies degree DEUG in English Studies at Cadi Ayyad University, Faculty of letters and human sciences in Marrakech, Morocco.
- BA degree in English linguistics at Cadi Ayyad University, Faculty of letters and human sciences in Marrakech, Morocco.

## **My motivations and interests evolve in identifying and consolidating the following issues:**

- Generate and publish my own Linguistic resources,
- Apply Linked Data principles and Semantic Web technologies (Ontologies, RDF, Linked Data) into the field of language resources.
- Use of the principal models used for representing Linguistic Linked Data, in particular OntoLex lemon,
- Learn about Linked Data-based NLP workflows and applications,
- Learn about potential benefits and applications of Linguistic Linked Data for specific use cases,
- Find answers for many different questions about the integration of my terminology resources according the W3C requirements and get the skills and knowledge required.

**Lucía Sánchez  
González**

***Ontology Engineering  
Group, Universidad  
Politécnica de  
Madrid***

# A little bit about myself

Bioinformatics

**Bachelor's  
Degree**

Biochemistry, Universidad Autónoma de Madrid



**Master's  
Degree**

Computational Biology, Universidad Politécnica de Madrid



**Master's  
thesis**

Spanish annotated corpus with biomedical entities and semantic relations

Natural Language Processing

Biomedical domain

**Elvin Huaman**

***University of  
Innsbruck, Austria***

# Elwin Huaman

## Background:

- PhD in Computer Science, 2019 (ongoing)  
Innsbruck University, Austria
- Msc in Web Engineering, 2017  
Oviedo University, Spain
- Msc in Digital Information Systems, 2014  
Salamanca University, Spain
- BSc in Systems Engineering, 2012  
National University of Altiplano, Peru

## Interest:

- Knowledge Graphs Curation: Assessment, Cleaning, and Enrichment
- Minority languages: Quechua, Aymara
- Semantic Web Technology
- Natural Language Processing
- QUIPU [Quechua Language Knowledge Graph] It can preserve as much as possible all the Quechua knowledge in a structured format.

**Martin Alejandro**  
**Chaya**  
***Universidad***  
***Politécnica de***  
***Madrid***



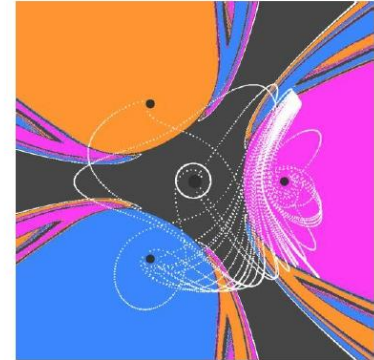
UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

Mathematics and Computer Science



FACILE  
Accesible Wikipedia

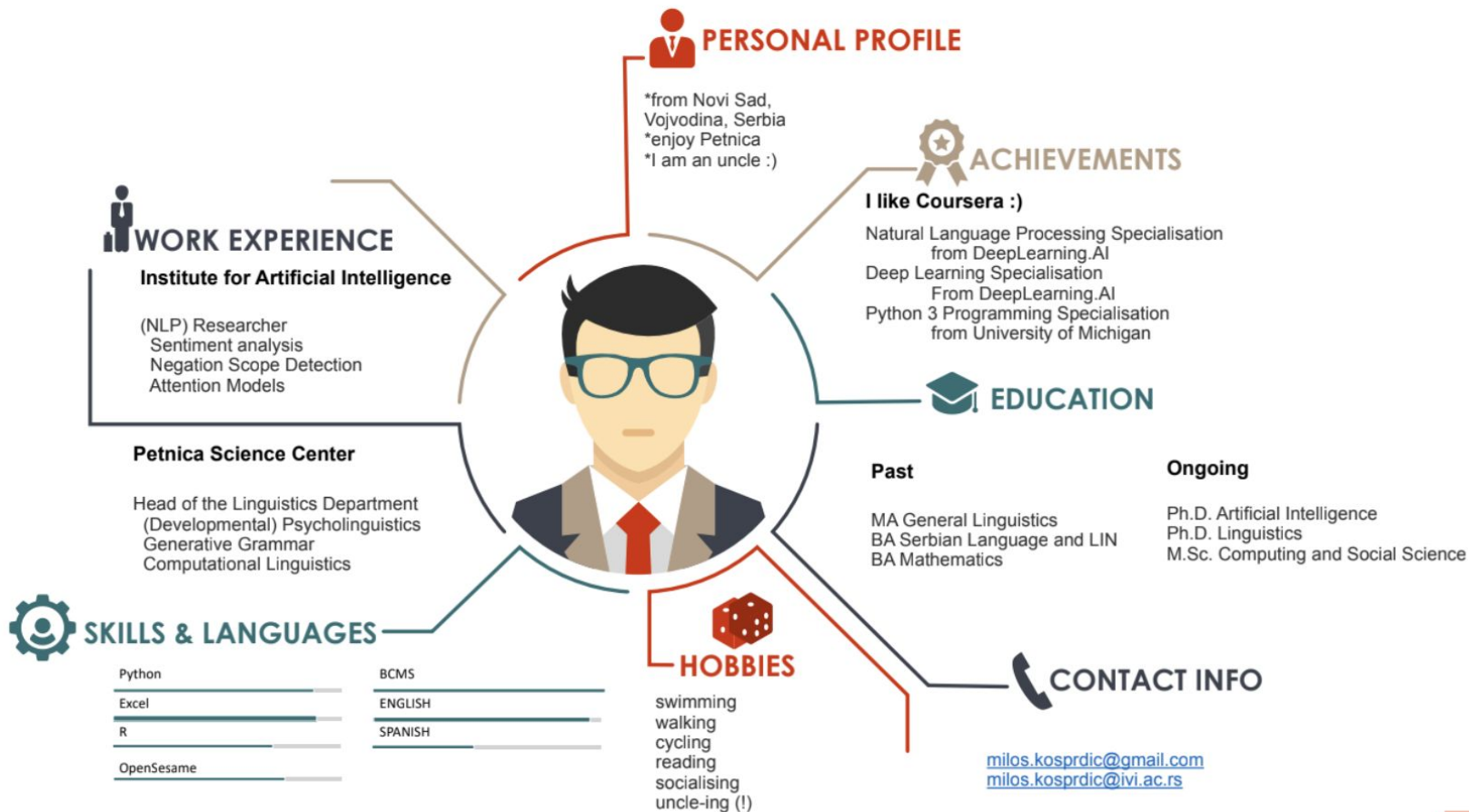
Hobbies





**Miloš Košprdić**  
***The Institute For***  
***Artificial Intelligence***  
***Research and***  
***Development of***  
***Serbia***

# About Me



**Tijana Radović**

***Laboratory for***

***Artificial***

***Intelligence,***

***University of***

***Belgrade***

# Tijana Radović

---

- PhD student at the Artificial Intelligence programme, University of Belgrade
- Working in Petnica Science Center
- Interested in linguistics, and NLP
- Currently developing lexicon of sentiments in Serbian language

**ZAIDA**

**BARTOLOME**

***Universidad de Las  
Palmas de Gran  
Canaria***



I am a civil engineer



I am a translator



My research:  
Specialised language of  
architecture and engineering  
in french and spanish

The language of architecture is one of the  
least analyzed specialized languages in  
French and Spanish



NEED

New lexicographic resource specialized in architecture,  
bilingual French-Spanish, digital and interconnected



My PhD thesis  
27/04!!



203 entries

101 French entries  
102 Spanish entries



XML File

TEI-LMF Format



**Valeria Caruso**  
***University of Naples***  
***'L'Orientale'***

# Valeria Caruso in URIs

DIZIONARIO MULTILINGUE DEL COVID-19 - COVID-19 MULTILINGUAL DICTIONARY - COVID-19 MEHRSPRACHIGES WÖRTERBUCH (2021)



UNIVERSITÀ DI NAPOLI  
L'ORIENTALE

## Procomplement Verbs

en avoir (FR), avercela (IT), cogerla (SP), 'to be angry with someone'

«they have a meaning of their own, which cannot be de-rived from the main verb, or which is very opaque...» (De Mauro, GRADIT, 1999)

## Idiomatica

<https://marvelapp.com/prototype/14e9g8cj>

Italian Idioms dictionary app by **Tap** prototype

<https://marvelapp.com/prototype/6g061f4>

Italian Idioms dictionary app by **Scroll** prototype

<https://www.unior.it> > didattica >

L'Orientale Apple Foundation Program – UniOr

L'Orientale Apple Foundation Program.

<https://www.lexonomy.eu/>

Dizionario Multilingue del Covid-19 - Lexonomy

<https://www.lexonomy.eu> > ...

Dizionario Multilingue del Covid-19 - Covid-19 Multilingual Dictionary - Covid-19 Mehrsprachiges Wörterbuch. sperimentale sviluppato durante il .

## Web Linguistics Resources

<http://www.weblinguisticresources.org>

A structured database of **linguistic resources** in the **web**. Sustained by academic grants from the University 'L'Orientale'. Developed by: Valeria Caruso ...

DIZIONARIO MULTILINGUE DEL COVID-19 - COVID-19 MULTILINGUAL DICTIONARY - COVID-19 MEHRSPRACHIGES WÖRTERBUCH (2021)

SEARCH

search



starts like this

Valeria Caruso Supervisione: Alessandra Chervino e Giulia Daniele

allentamento relaxation (eng) Lockerung (deu) Relajación (esp) afrouxamento (pt) ослабление (ru)

allentare ease (eng) lockern (deu) allentare afrouxar (pt) снять (ru)

ammalarsi get sick (eng) заболеть (ru) adoecer (pt)

asintomatico asymptomatic (eng) ohne fofei (deu) asymptomatique (fr) asintomático (esp) assintomático (pt) бессимптомный (ru)

bollettino quotidiano daily bulletin (fr) chiffre quotidien (fr) Ежедневный бюллетень (ru)

antidote (eng) антидот (ru)

antigene antigen (eng) مستضد (ar) antigeno (esp) антиген (ru)

antibody cocktail (eng)

caso sospetto suspected case (eng) Verdachtsfall (deu) caso sospechoso (esp)

contagio infection (eng) Ansteckung (deu)

contaminazione contamination (fr) besmetting (nl) заражение (ru) contagio (esn) bulasma (tr)

cluster cluster (eng) cluster (fr)

contenere contain (eng) eindämmen (deu)



**Erik Körner**  
***Saxon Academy of  
Sciences and  
Humanities in Leipzig***



UNIVERSITÄT  
LEIPZIG

Bauhaus-Universität Weimar



Sächsische Akademie der Wissenschaften  
zu Leipzig



CORPORA COLLECTION  
LEIPZIG



Crawling  
Underresourced Languages

DWEE



**Isam Diab**  
***Universidad***  
***Politécnica de***  
***Madrid***



Universidad de Valladolid

Spanish Language and Literature  
2014-2018



UNIVERSITAT DE  
BARCELONA

Master in Cognitive Science  
and Language  
2018-2019



*emagister:*



expert.ai

Computational Linguist  
2018-2020

Cognitive  
Linguistics

+

Artificial  
Intelligence

=

Cognitive  
Accessibility

