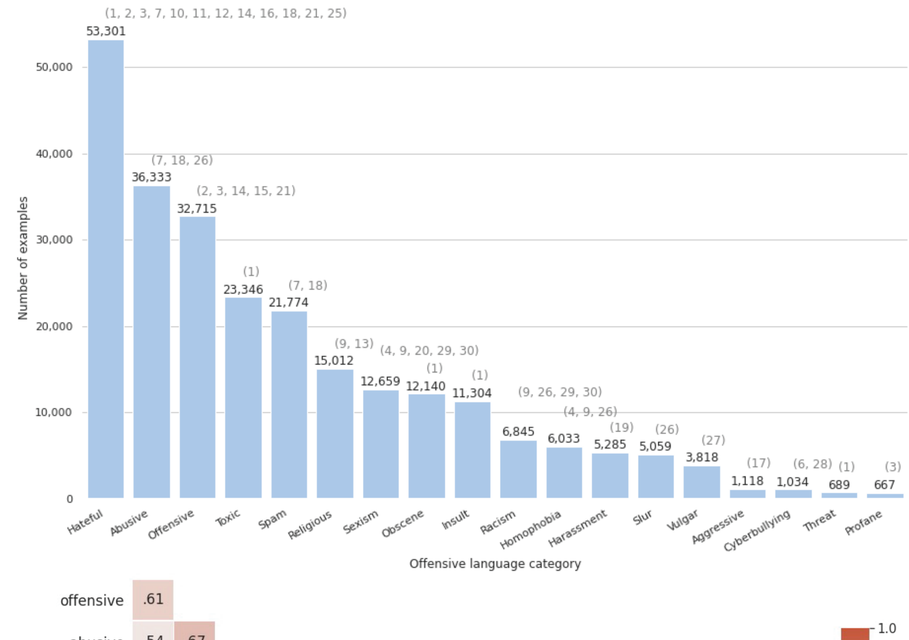
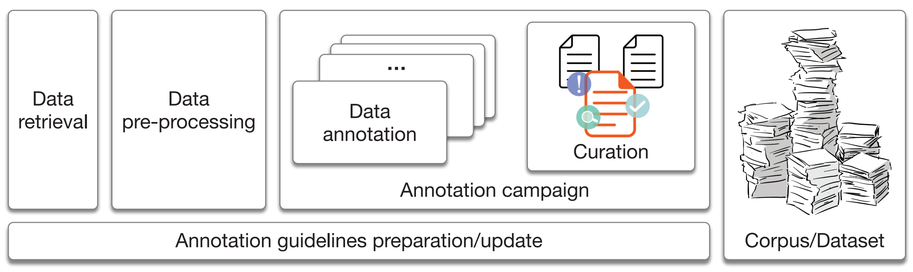
UC 4.1.2 Offensive Language Categorization Gold Standard for the LLOD Schema

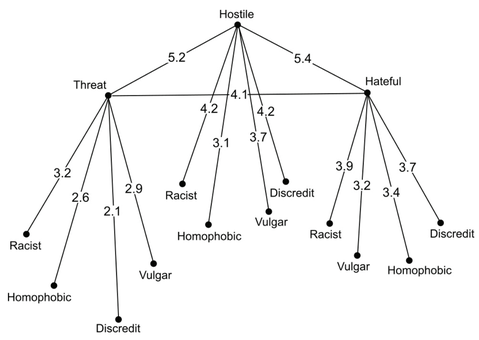
Offensive language datasets

- Over 60 datasets identified
- 25 datasets retrieved and imported
  - 18 offensive language category groups
  - various sources (e.g., Twitter, forums, reddit, comments)
  - 834,127 (731,936 unique) examples
- Problems
  - vague offensive language term definitions / no guidelines
  - incompatible data for joint analysis

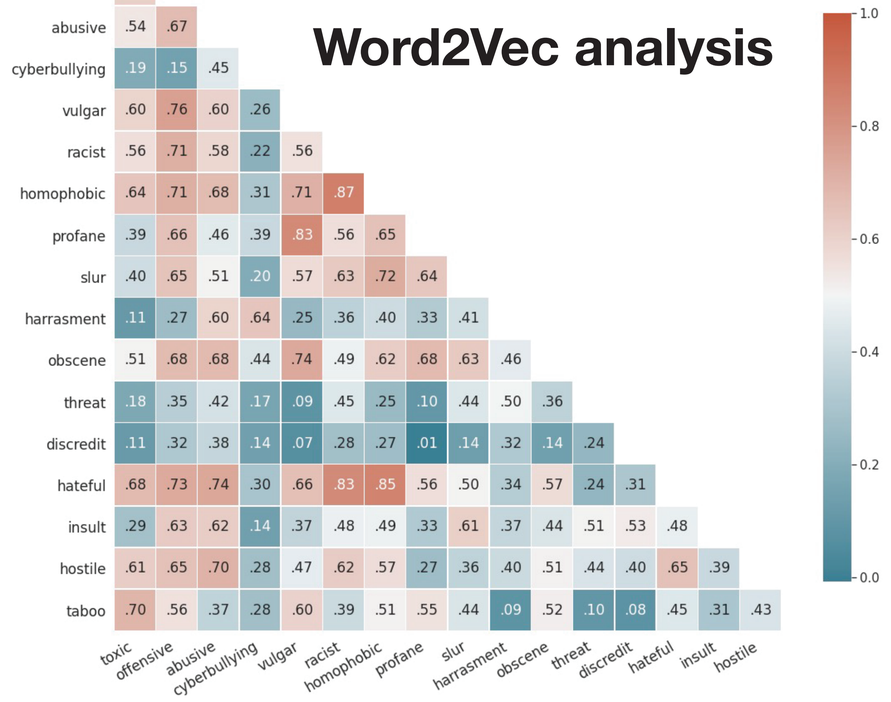


Dataset annotation process + Survey analysis

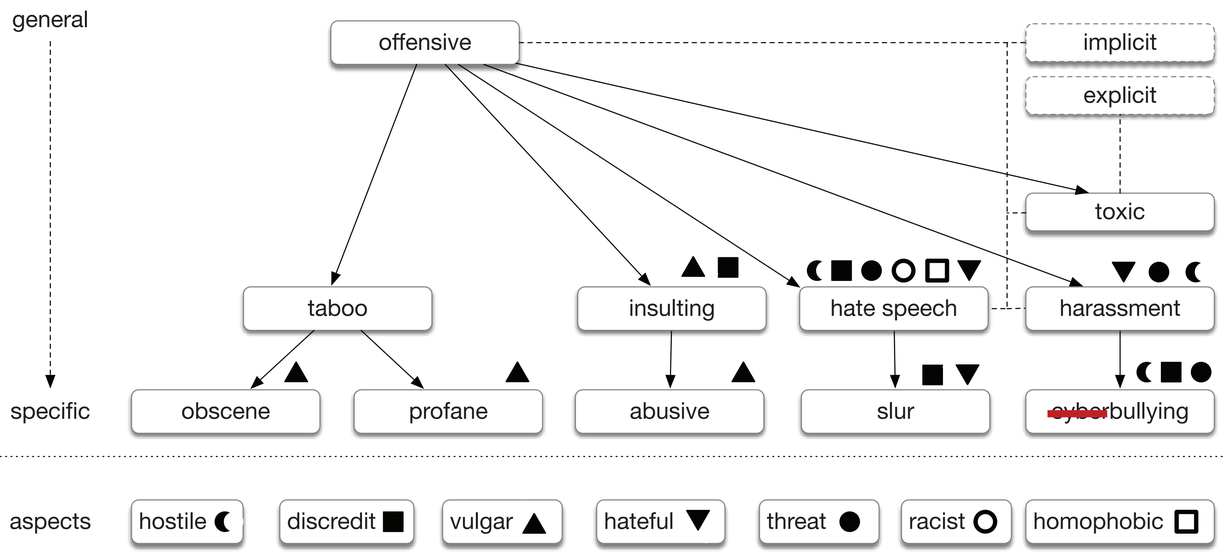




Word2Vec analysis



First schema proposal (Skopje, 2021)



Simplified version proposal (2022) and validation with annotation

- **SIMPLIFIED OFFENSIVE LANGUAGE TAXONOMY SOL Taxonomy (Lewandowska-Tomaszczyk 2022)**

1. **OFFENSIVE [YES or NO]**
2. **Target 1 Individual // Group // Ind wrt Gr/Gr wrt Ind [by REFERENCE TO STEREOTYPES]/non-targeted**
3. **Target 2 : present//absent**
4. **Vulgar [YES OR NO]**
5. **Choose either (i) OR (ii); Then select (iii) OR (iv) or BOTH (iii) AND (iv)**

(i) HATE SPEECH [individual or group; offense by REFERENCE TO GROUP STEREOTYPES]

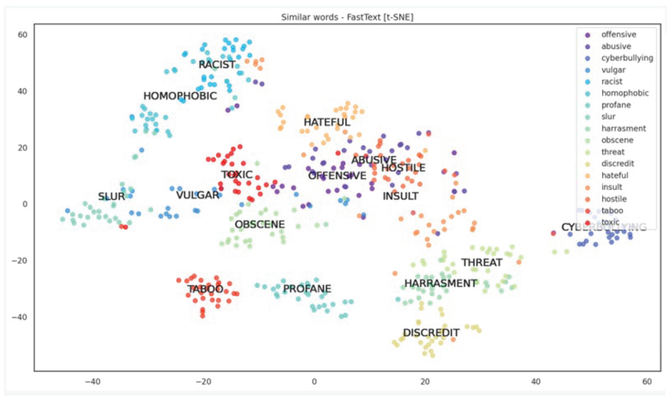
(ii) INSULT [addressed to: individual or group - varied offense types but NOT by group stereotypes]

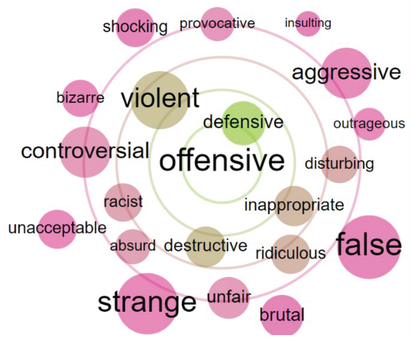
(iii) DISCREDIT [individual or group//on various grounds – lying-cheating, immorality, unprofessionalism, unfairness]

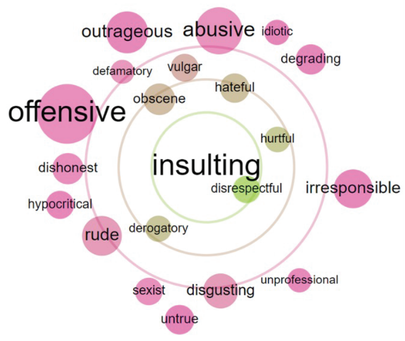
(iv)THREAT [individual or group, inducing fear]

6. Aspects – [Choose one or more] *[racist] [xenophobic] [homophobic] [sexist] [profane (religion)] [ageism] [physical/mental disabilities [ableism]] [social class [classism]] [ideologism] [other]*

7. Select categories below – [Choose one or more] RHETORICAL QUESTIONS / METAPHOR/SIMILE / IRONY / EXAGGERATION / OTHER



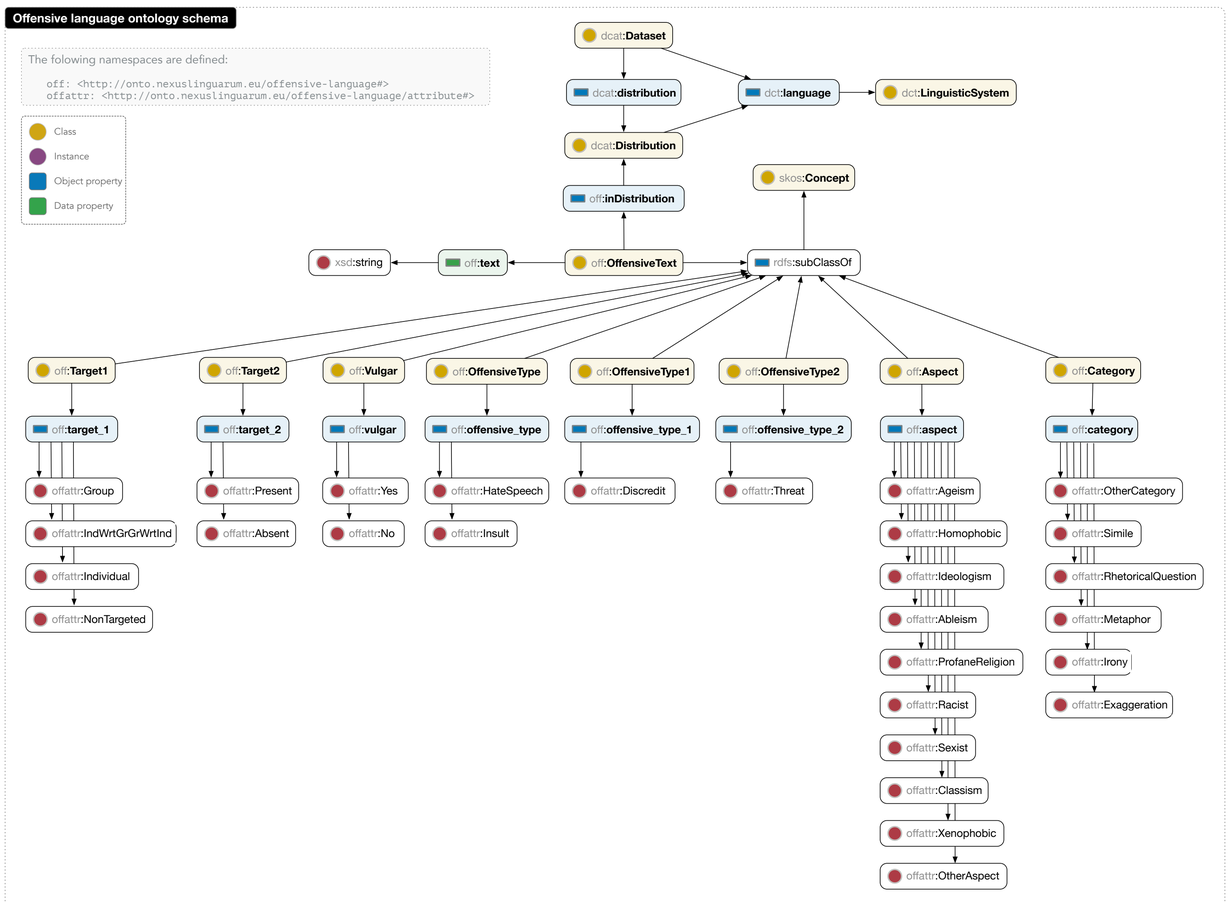




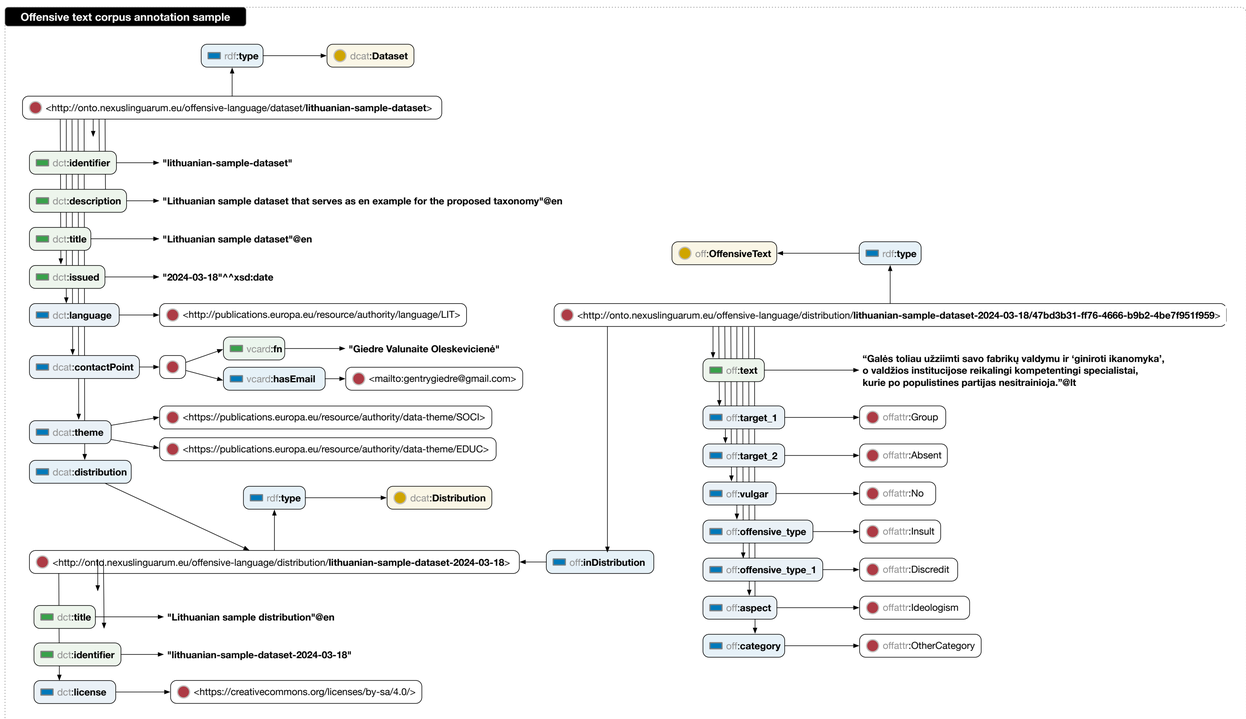
Final simplified offensive language schema (Athens, 2024)

Representative samples in **7 languages**: Lithuanian, Croatian, Czech, English, Hebrew, Kazakh and Polish

Offensive language ontology schema



Offensive text corpus annotation sample



https://github.com/szitnik/offensive-language-ontology

Barbara Lewandowska-Tomaszczyk, Slavko Žitnik, Chaya Liebeskind, Giedre Valunaite Oleskevicienė, Anna Bączkowska, Paul Wilson, Marcin Trojszczak, Ivana Brač, Lobel Filipić, Ana Ostroški Anić, Olga Dontcheva-Navratilova, Agnieszka Borowiak, Kristina Despot, Jelena Mitrović and others …