TZOS: an Online Terminology Database Aimed at Working on Basque Academic Terminology Collaboratively

http://tzos.ehu.es

Izaskun Aldezabal (izaskun.aldezabal@ehu.eus)
Jose Maria Arriola (josemaria.arriola@ehu.eus)
Arantxa Otegi (arantxa.otegi@ehu.eus)

LREC 2022, Marselle, 13th edition, January 20-25

Framework: Communicative Theory of Terminology (CTT)

- Terms are activations of different semantic features of a lexical item in a context of use (Cabré, 2001)
- Specialized discourse community -> natural "context of use" for natural development and self-regulation
- Members of these comunities and their texts - > experts and appropiate corpora

  Experts (cognitive and pragmatic knowledge) = authors (of the texts) = users (of the term) = contributors (for TZOS)

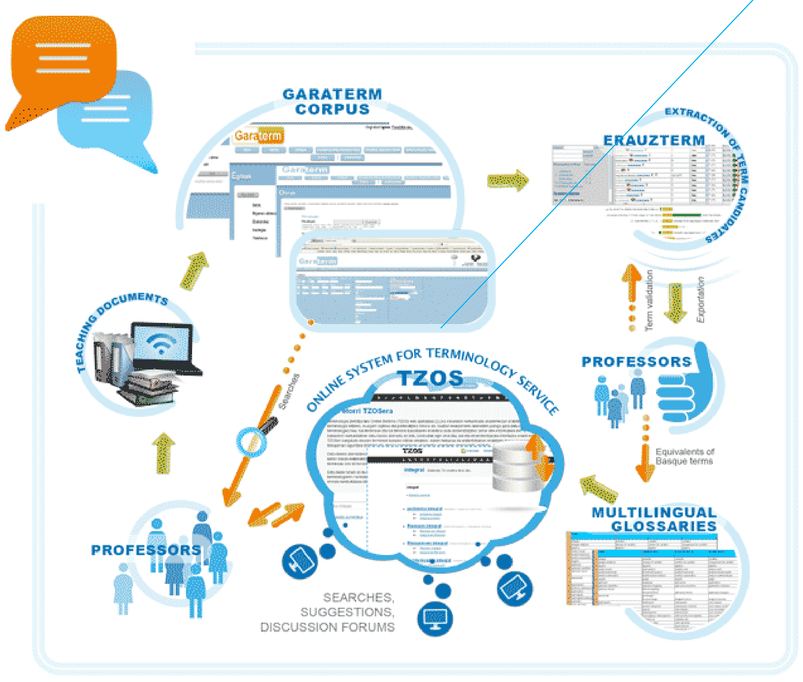
- An *active* description -> users involved in the description of terms
- Less dependent on other languages
  - **Basque**: in the process of development of terminological elaboration -> university = agent of revitalization

GARATERM environment (Zabala et al., 2013)

- **GARATERM**: academic corpus **related to a specific subject** (exams, exercises, practical works, notes, presentation slides…)
- **Erauzterm** (Gurrutxaga et al. 2005): automatic term extractor. Term candidates from the texts of each expert. Offered as a list in the GARATERM platform for been validated and translated into other languages: **multilingual term glossaries**
- **TZOS**: the online terminology database where all these **terms are stored**

GARATERM environment

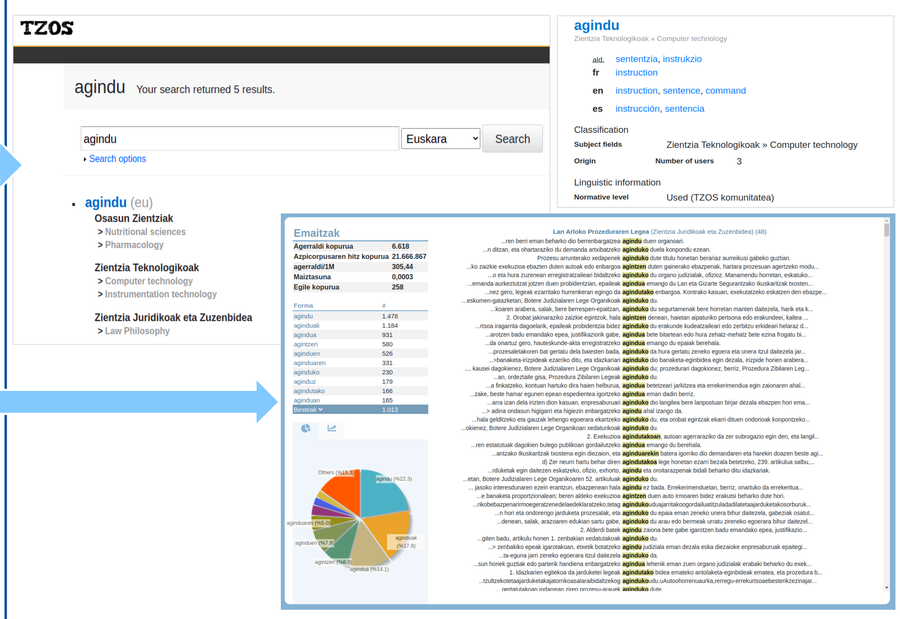
21,762,846 words
839 experts

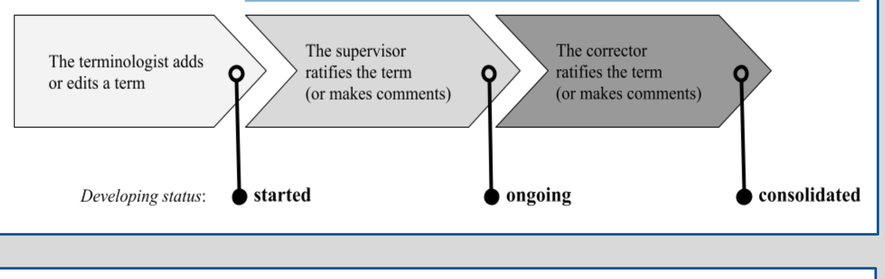


TZOS (Arregi et al., 2013)

- **Terms in TZOS**
  - Classified according to Unesco fields and subfields
  - Term-entry: one entry per subfield
    - Variants, definitions…
    - Usage (origin)
  - Collision: if equivalent terms
- **Connection with GARATERM Corpus**
  - By clicking on a term in Basque the contexts in which this term is found in GARATERM is shown
- **User Profiles and Term Development**
  - 3 user profiles
  - 3 developing status

136,540 term-entries
180 experts



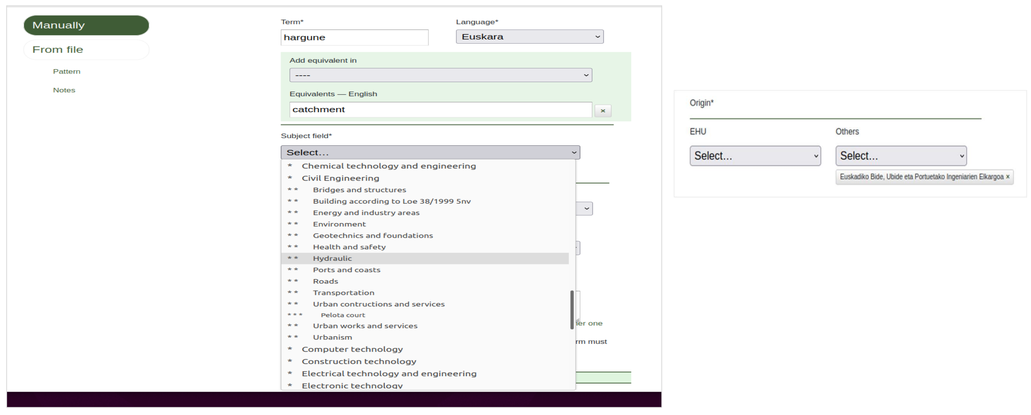


IZIBI-TZOS: a Civil Engineering oriented TZOS (izibi-tzos.ehu.eus/)

- **Civil Engineering area**
  - Low-sourced in Basque
  - The CTT approach needed for the description
  - An agreement between the UPV/EHU and the College of Civil, Channels and Ports Engineers of the Basque Country (involved in a Basquisation plan)
- **Distinctive features**
  - Inclusion of subfields within the Civil Engineering subfield
  - New origins from the professional area

82,359 term-entries

- From TZOS:
  - Civil Engineering subfield
  - Civil Engineering degree
- Currently:
  - Enriching it by experts from both the academic and professional areas



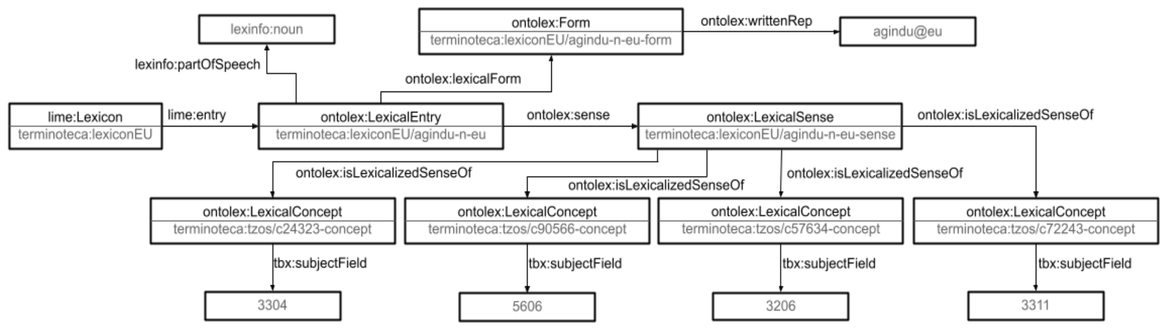
TZOS as Linguistic Linked Open Data

publicly available in
https://www.ixa.eus/node/12983

Prototype for a Resource Description Framework (**RDF**) representation based on ***Ontolex-lemon* model** (Cyganiak et al., 2014)

- Main elements of lexical descriptions: lexical entries, lexical concepts, senses
- Other information: translations and variants (linking lexical senses from the same lexicon denoting the same concept)

Lexical-entry: *lime:Lexicon* (Basque, Spanish, English, French, and Latin term lists; extra lexicon for symbols) *ontolex:LexicalConcept* class (concepts that lexical entries evoke) *ontolex:LexicalSense* intermediate class (sense of the lexical entry), *tbx:subjectField* (subfield in which a concept is used); URI numeric identifier (merging terms from different terminologies at the lexical entry-level). e.g.: lexiconEU/agindu-n-eu; lexiconEN/command-n-en



Conclusions

(1) To share terminology resources and facilitating discussion groups to compensate for the lack of fluid communication networks in Basque among experts/teachers
(2) To disseminate academic terminology to be taken into account by normalizing institutions
(3) To monitor texts and terminology used in academic communication in Basque
(4) To open a new line of work, extending TZOS to the Civil Engineering professional area
(5) To interlink TZOS to other resources in the LLOD

Future challenges:

(1) Incorporation and codification of semantic tags or classes to enrich the term-entry
(2) To extend the collection and processing of texts necessary to make the corpus GARATERM representative
(3) To integrate the results of TZOS and GARATERM with the agents and entities responsible for the normalization of Basque

References

Arregi, X., Arruarte, A., Artola, X., Zabala, I., & Lersundi, M. (2013). TZOS: An On-Line System for Terminology Service. In *Actualizaciones en Comunicación Social. Actas XIII Simposio Int. de Comunicación Social, Santiago de Cuba*, pages 400-404. Centro de Lingüística Aplicada.
Cabré, M.T. (2001). Sumario de principios que configuran la nueva propuesta teórica. In M.T. Cabré & J. Feliu (Eds.), La terminología científico-técnica, IULA, Universitat Pompeu Fabra, pages 17-26.
Cyganiak, R., Wood, D., Lanthaler, M. (2014). Resource Description Framework (RDF) 1.1: Concepts and Abstract Syntax: W3C Recommendation 25 February 2014.
Gurrutxaga, A., Saralegi, X,, Ugartetxea, S., & Alegria, I. (2005). Eraucterm: euskarazko terminoak eraucteko tresna erdiautomatikoa. Mendebalde Kultur Alkartea, IX. Jardunaldiak: Euskera zientifiko-teknikoa.
Zabala, I., Lersundi, M., Leturia, I., Manterola, I., and Santander, G. (2013). GARATERM: euskararen erregistro akademikoen garapenaren ikerketarako lan-ingurunea. In Ugarteburu terminologia jardunaldiak (V). terminologia naturala eta terminologia planifikatua euskararen normalizazioari begira, pages 98-114.