

DC4 Punsisi LIYANAGE

IRP4: Stereotyped language related to gender, race and ethnicity in digital dictionaries and thesauri

Research fields: Applied Linguistics, Gender Linguistics, Contrastive Linguistics

Co-supervisors of IRP4:

Hélène Vinckel-Roisin, helene.vinckel-roisin@univ-lorraine.fr

Mirjam Schmuck, mirjam.schmuck@hum.ku.dk

This PhD position is part of a double-degree programme. Upon completion, candidates will be awarded **one joint doctoral degree**, on two PhD diplomas, following research stays at both institutions: **Université de Lorraine** (diploma conferred by **Université de Lorraine**) and the **University of Copenhagen** (diploma conferred by the **University of Copenhagen**).

Scientific context of the Individual Research Project IRP4:

IRP4 is embedded in **Work Package 1 (WP1)** ‘Multilingual communication and resources in the Digital Age: discourses, practices, stereotypes’. WP1 investigates how the concepts of Language Awareness and pluri-/multilingualism are discursively constructed, ideologically framed, and technologically mediated in digital contexts across various societal fields (e.g., media, education, industry, politics). It will provide a research-based catalogue of arguments and language ideologies, with particular attention to issues of linguistic justice, gender fairness, and inclusion.

In this context, IRP4 focuses on identifying and analysing stereotypical language patterns and existing biases —particularly related to gender, race and ethnicity—in digital dictionaries and thesauri.

Recruiting University: Université de Lorraine (UL), Nancy, France

Partner University for the Joint Doctorate: University of Copenhagen (UCPH), Denmark

Objectives:

In the 21st century, stereotypical language is still widespread, accumulated in dictionaries, reinforcing stereotypes in easy online access. IRP4 aims therefore to:

- 1) analyse the representation of gender, race and ethnicity in digital (corpus-based) mono-/multilingual dictionaries and user interfaces (e.g. word thesaurus) to uncover fairness harms and discrimination against diverse demographic groups in lexical entries in terms of definitions, “typical” examples, collocations and synonyms;

2) raise users' and developers' Critical Language Awareness of stereotypical representations arising, among others, from biased internal datasets, e.g. digital corpora mainly relying on newspaper texts (collocations such as woman+raped/abused; Syrian+criminal/terrorist/arrested);

3) disclose outdated role models that are perpetuated in new technologies;

4) establish a set of measures to confront biases and stereotypical representations in digital resources aimed at civil society, e.g., by optimising underlying datasets with associated partners.

Expected Results:

- A comprehensive picture of gender, race and ethnicity's stereotypes rooted in digital mono- and multilingual dictionaries;
- List of data sources that allow a non-biased representation;
- Reduce fairness harms in widely used online dictionaries and user-interfaces

Secondment:

- Berlin-Brandenburg Academy of Sciences and Humanities, Berlin (Germany), 4months

Languages: English, German, French