Multilingual Contextual Affective Analysis of LGBT People Portrayals in Wikipedia

Chan Young Park*, Xinru Yan*, Anjalie Field*, Yulia Tsvetkov

Carnegie Mellon University
Language Technologies Institute

GOAL: Analyze how LGBT people are portrayed on English, Russian, and Spanish Wikipedia
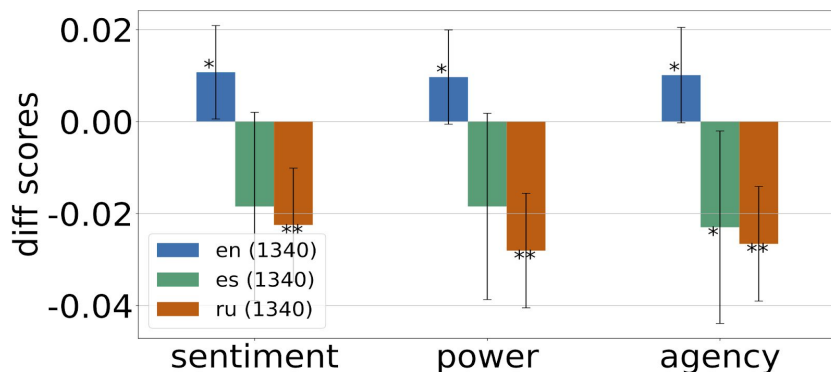
> **English Wikipedia:**
> He *accepted* the option of injections of what was then called stilboestrol.
>
> **Spanish Wikipedia:**
> Finalmente escogió las inyecciones de estrógenos.
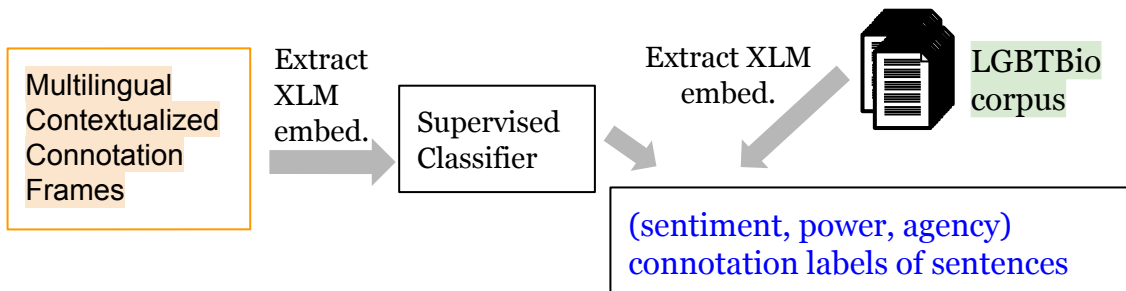> *Finally he chose estrogen injections.*
>
> **Russian Wikipedia:**
> Учёный предпочёл инъекции стильбэстрола
> *The scientist preferred stilbestrol injections.*

METHOD: 2 Datasets + Multilingual Model

1. Collect a multilingual corpus of verbs annotated for sentiment, power, and agency connotations. → Multilingual Contextualized Connotation Frames
2. Collect Wikipedia bio articles of 1.3k LGBT people → LGBTBio Corpus
3. Use a machine learning model to infer connotations of verbs in Wikipedia articles about LGBT people and non-LGBT people

Multilingual Contextualized Connotation Frames → Extract XLM embed. → Supervised Classifier → Extract XLM embed. → LGBTBio corpus → (sentiment, power, agency) connotation labels of sentences

FINDINGS:

- Connotations about LGBT people are more negative than connotations about non-LGBT in Russian, but not in English or Spanish
- checkout our paper for additional analyses!