

Measurement of Qualitative Gender Presence in News

Proposer(s) / Proposatzailea(k):

Elhuyar Fundazioa

Contact / Kontaktua:

Xabier Saralegi, Text Analytics manager (x.saralegi@elhuyar.eus)

Description / Deskribapena

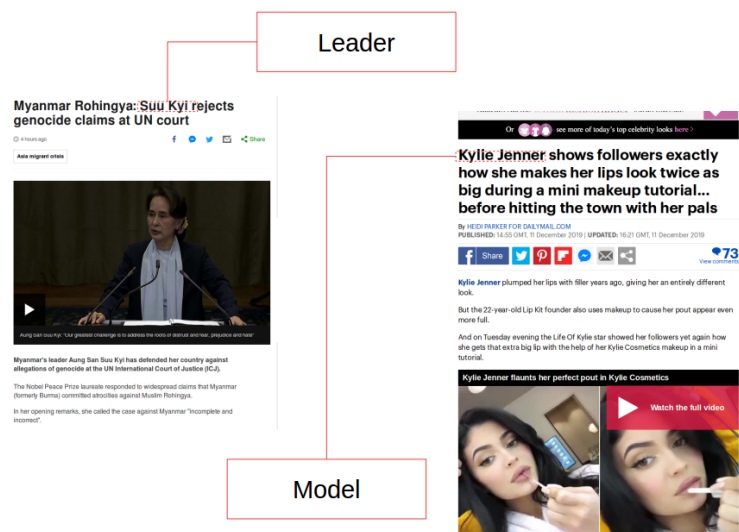
The press offers readers a macro-story based on its interpretation of reality. This macro-story consists of a multitude of micro-stories starring in most cases by people.

From a gender perspective it is of great interest to analyze the differences of prominence between the different genres in the macro-story. This type of analysis allows to detect both gender biases in the interpretation and unbalanced realities. The function of these analyzes is to be able to make precise diagnoses in order to design effective corrective measures.

This type of analysis is usually done manually on news samples (Macharia et al., 2010; Peña-Fernández & Martínez, 2017). The manual analysis of the texts allows measurements of both quantitative and qualitative presence.

In recent years, analyzes based on automatic methods have been proposed (Jia et al., 2016; Lansdall-Welfare et al., 2017; Saralegi, 2019), primarily on natural language processing methods. However, these methods are oriented to a quantitative measurement of the presence of each gender according to different variables. Its main lack, therefore, is not being able to offer qualitative presence measurements.

In this work we propose to solve this lack and develop techniques to measure the qualitative presence by detecting roles in the news.



- Jia, S. et al. Women are seen more than heard in online newspapers, 2016.
- Lansdall-Welfare, T. et al. Content analysis of 150 years of British periodicals, 2017.
- Macharia, Sarah, Dermot O'Connor, and Lilian Ndangam. Who makes the news?: Global media monitoring project 2010. World Association for Christian Communication, 2010.
- Peña-Fernández & Martínez. Análisis de la presencia de las mujeres en los medios de comunicación vascos 2015-2016, 2017.
- Saralegi, Xabier. Euskal Herriko prentsa digitalean emakumezkoek zenbateko presentzia duten kuantifikatzen testuen meatzaritz baliatuta, 2019.

Goals / Helburuak

Main goal:

- Develop technique for measuring qualitative gender presence in news.

Specific goals:

- Use of NERC models to identify people and their genders in news.
- Development of techniques for identifying people's roles in news.

Requirements / Betebeharrak

Computer programming skills will be needed, preferably Python. Experience in training and using neural language models for text classification would be interesting. The working environment will be Linux.

Tasks and plan / Atazak eta plana

- T1: Adapt NERC models for identification of person names and their genders (Gender name dictionary will be provided by Elhuyar).
- T2: Definition of role identification task.
- T3: Dataset creation for the role identification task (news corpus will be provided by Elhuyar).
- T4: Implementation of supervised classifier for role identification:
 - Bag of words + Logistic Regression/SVM.
 - Static embeddings + Logistic Regression/SVM.
 - Static embeddings + RNN.
 - Neural Language Model + fine tuning or feature based.