Women's Representation in Cartoons

Charlotte Weirich, Alex Valisu

Project Description

The project focuses on analyzing the frequency at which women appear in English-speaking media and fiction, specifically utilizing the Bechdel-Wallace test: "Are there at least two women who talk with each other about something besides a man?" In addition, we are looking into expanding our data by computing the ratio of tokens spoken by women and tokens spoken by other genders, or comparing the amount of time female characters speak compared to other speakers.

We are focusing our project on gathering data from cartoons primarily aimed at a child or young adult audience, with the starting year of 2000 or later. Our corpus will be composed of only English data, sourced from either the original release or from the official or widely accepted subtitles in the case of foreign cartoons.

The first step is to create our own corpus, since corpora available online don't tend to contain speaker data, which are crucial to our project. We plan on gathering subtitle data from online sources, then manually annotating speaker data, and in addition manually or semi-automatically annotating audience and topic data. We will also come up with more possible annotation layers to get a more detailed account of women's representation in media.

Work Distribution

Charlotte Weirich:
- coming up with annotations and how to best implement them
- implement the automatic systems for named-entity recognition, coreference resolution, topic annotations, speaker ratio by gender

Alex Valisu:
- scraping the internet for subtitle files in the .srt format to add to the corpus
- manual annotation of speaker data, manual correction of automatic system annotation