Combining Cultural Educational Theory and Corpus Linguistics

Anna Moskvina
Universität Hildesheim, Institut für Informationswissenschaft und Sprachtechnologie.
E-mail: moskvina@uni-hildesheim.de

- This study is carried out in the framework of the BMBF-funded project Rez@Kultur.
- Aim: the analysis of online reviews of cultural artefacts = books, works of art, art exhibitions:
  1. components
  2. structure
  3. reception process
  4. production process
  5. conditions influencing these processes

- Current RESEARCH GOAL:
  - Extraction of textual evidence for cultural education processes:
    1. Marotzekji 1990:
      \[ \Rightarrow \] processes where the conception which an individual has of himself or herself, and/or his/her conception of others and of the world around them undergo a change
    2. Reinwand-Weiβ 2012:
      \[ \Rightarrow \] processes take place in the subjects themselves, when they decide, because of their interests, to get involved with an artefact
    - Examples:
      1. "[...]. . . und dieses Buch hat mir die Augen geöffnet" ("[...] and this book was an eye-opener for me") \[ \Rightarrow \] Change of Perspective
      2. "[...]. . . bekommt man eine ganz andere Sichtweise auf solche Themen" ("[but] because of [...] you get a totally different perspective onto such topics") \[ \Rightarrow \] Change of Perspective
      3. "Ein Buch für alle, die [...] unsere emotionalen Reaktionen besser verstehen wollen" ("A book for everyone who wants to better understand [...] our emotional reactions") \[ \Rightarrow \] Voluntary Act
      4. "Auch wenn nicht jeder Baum beschrieben wird [...] das ich mir alles sehr gut vorstellen könnte" ("Even if not every tree is described [...] I could imagine everything") \[ \Rightarrow \] Change of Perspective

- RESEARCH WORKFLOW:
  1. Manual annotation of components of reviews, their contents, reviewing processes, the emotions of reviewers, their recommendations etc. \[ \Rightarrow \] 430 reviews
  2. Analysis of sentences with 4 types of labels
  3. Manual extraction of indicators
  4. Application of indicators in queries to the combined corpus of online book reviews from Amazon and other platforms
  5. Use of found sentences as training data for a classifier

- EXAMPLE PATTERNS FOR CATEGORY A5
  (Writing about the education effect)
  1. Pattern 1: "vertiefen" + indirect object

- Results of TRAINING
  - SVM model, A5 vs all, simple bag-of-words
  - initial annotated data set
    test: 30, train: 93
  - new annotated data set:
    test: 234, train: 1899

- FUTURE WORK
  1. Add corpora from other sources
  2. Add features to the learning algorithm (system of weights to tackle imbalance)
  3. Compare algorithm with human annotation (previously unseen texts)
  4. Repeat