Odeuropa

Analysis of Smell in Texts

Text Analysis

Consider data properties

- Different data sources: news, fairytales, books, historic texts,…

- Natural language of the text: English, Italian, French, Slovenian, …

- Importance of time in the data collection: date of news article, time period of a book,…

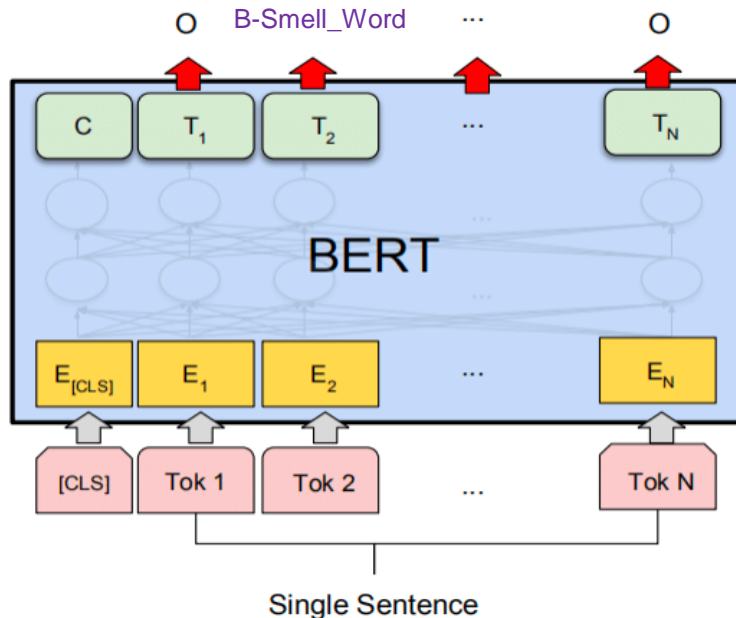Machine learning to extract olfactory information from texts written in the different languages

- Which contexts, places and events are mentioned in relation to olfactory information

- Which terminology is used

- How the information changes over time and across languages

Odeuropa text processing system

Machine learning to label each token in a sentence

The scent of flowers is often cloying

O   B-Smell_Word   B-Smell_Source   I-Smell_Source   O   O   B-Quality

O   B-Smell_Word   ...   O

Classes:
- Smell Source
- Quality
- Smell Word

Odeuropa text processing system

One morning the Artist had paused a moment to make a rough sketch of a man who [...] was selling **segments of round flat cheeses of goat's milk**; **vile smelling** compounds that, judged from their outer coating of withered leaves, straw, and dirt, would appear to have been made in a stable and dried on a rubbish heap .

smelling	segments of round flat cheeses of goat's milk	vile

 In its pure state, **civet** has, to nearly all persons, a **most disgusting odor**; but when diluted to an infinitesimal portion, its **perfume** is **agreeable** .

odor perfume	civet	most disgusting agreeable

Multilingual emotion recognition

Data: REMAN data with annotation of emotions with the roles of experiencer, cause, and target, 1720 sentence triples from Project Gutenberg www.gutenberg.org

Emotions

- Joy – glad, happy, ….
- Sadness – sad, broken-hearted,…
- Anger – furious, angry with,…
- Fear – afraid of, frightened,…
- Surprise – astonished, surprised,…
- Anticipation – looking forward, expecting,…
- Trust – trust,…
- Disgust – hate, abominate, cannot stand,…
- Other – any other emotion

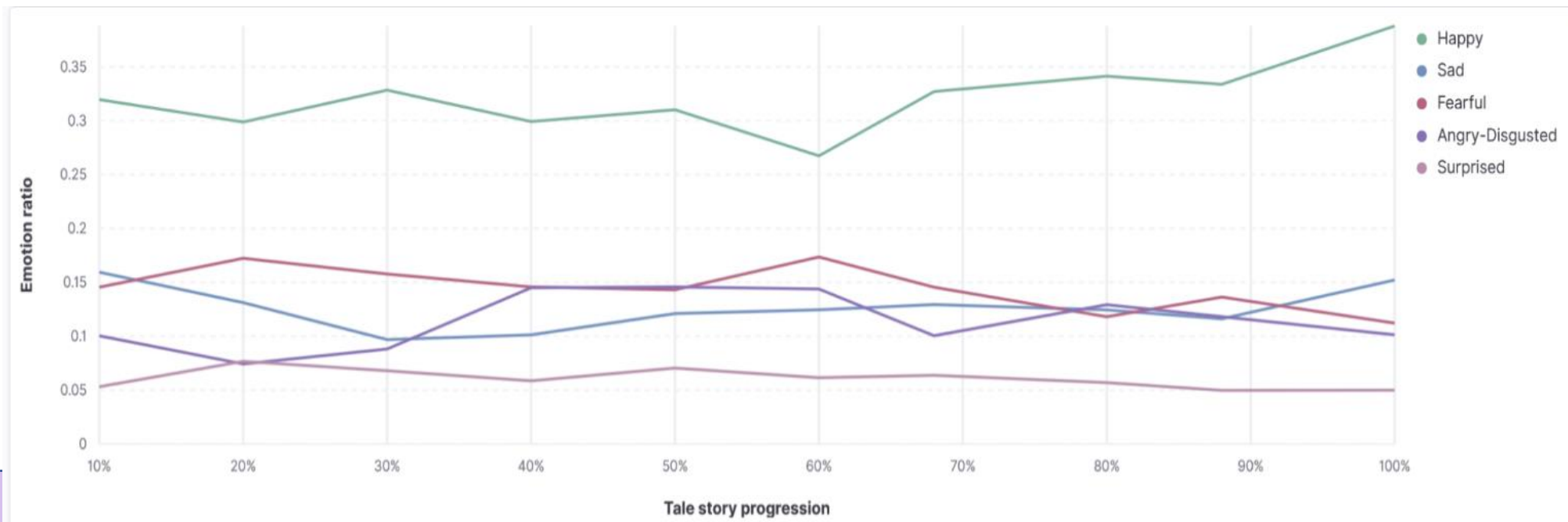Machine learning: XLM-Roberta, BERT

Multilingual emotion recognition

Fairy tales data

- Affect data, 1207 sentences: Angry-Disgusted, Fearful, Happy, Sad, Surprised
- Andersen tales and tales of Brothers Grimm, 33000 sentences

Analysis of emotions

- The most common emotion in fairy tale as a story progresses

Metadata for the source

Text items (`crm:E33_Linguistic_Object`) and Image items (`crm:E36_Visual_Item`) should be represented with the following metadata

- Author - `schema:autor`
- Title - `rdfs:label`
- Publishing date - `schema:dateCreated`
- Publishing place - `schema:locationCreated`
- Language - `schema:inLanguage`
- Genre - `schema:genre`
- Topic - `schema:about`
- Museum / library - `ecrm:P53_has_former_or_current_location`
- External URL - `schema:url`
- ISSN id - `schema:issn`