

Donnerstag  
09.03.2017  
10:30 – 11:15  
B4 1, Foyer

---

## Color distributions in German poetry

---

Laura Bostan  
*Trento University*  
[sarnthil@gmail.com](mailto:sarnthil@gmail.com)

Jonathan Oberländer  
*Trento University*  
[jonathan.oberlaender@gmail.com](mailto:jonathan.oberlaender@gmail.com)

We investigate the usefulness of color frequencies in identifying the period in which a poetic work was created. We predict that these features alone aren't able to discriminate time periods with great accuracy, but by adding them a classifier can get better results.

McManus (1983) showed that the distribution of color words in English and Chinese poetry correlates with the order of evolution of color words across languages (Berlin and Kay, 1963). Later, McManus (1997) continued the experiment computationally on a larger English dataset. We also intend to repeat this study for German poetry, and possibly for other languages.

We crawled all poems available on Project Gutenberg-DE<sup>1</sup>. The number of different authors was 299, most of which are German, Austrian or Swiss. The rest are of different nationalities (with their works translated to German). After the crawling, we counted the color word distribution for each author using a set of regular expressions.

We then balanced the raw counts by dividing the occurrence count of a color to the count of all word occurrences for each author. Using Wikipedia, the list of authors was automatically enriched with information about their years of birth and death. We intend to further enrich it with a label of literary period. As we were interested in seeing whether the color distribution changes over time, we created a small visualization<sup>2</sup> to explore the data, but no obvious development can be seen.

We are currently working on extending the experiments to include new datasets in languages other than German. Here, interesting future work would be to investigate the relationship between poetry in different source languages and different literary currents.

**References:** • McManus, I. C. (1983): Basic colour terms. *Language and Speech* 26, 247–252. • McManus, I. C. (1997): Half-a-million basic colour words: Berlin and Kay and the usage of colour words in literature and science. *Perception* 26, 367–370. • Berlin B., Kay P. (1969): *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA.

---

<sup>1</sup><http://gutenberg.spiegel.de/genre/gedicht-poem>

<sup>2</sup><https://jsfiddle.net/4w6vxlls/1/embedded/result/>