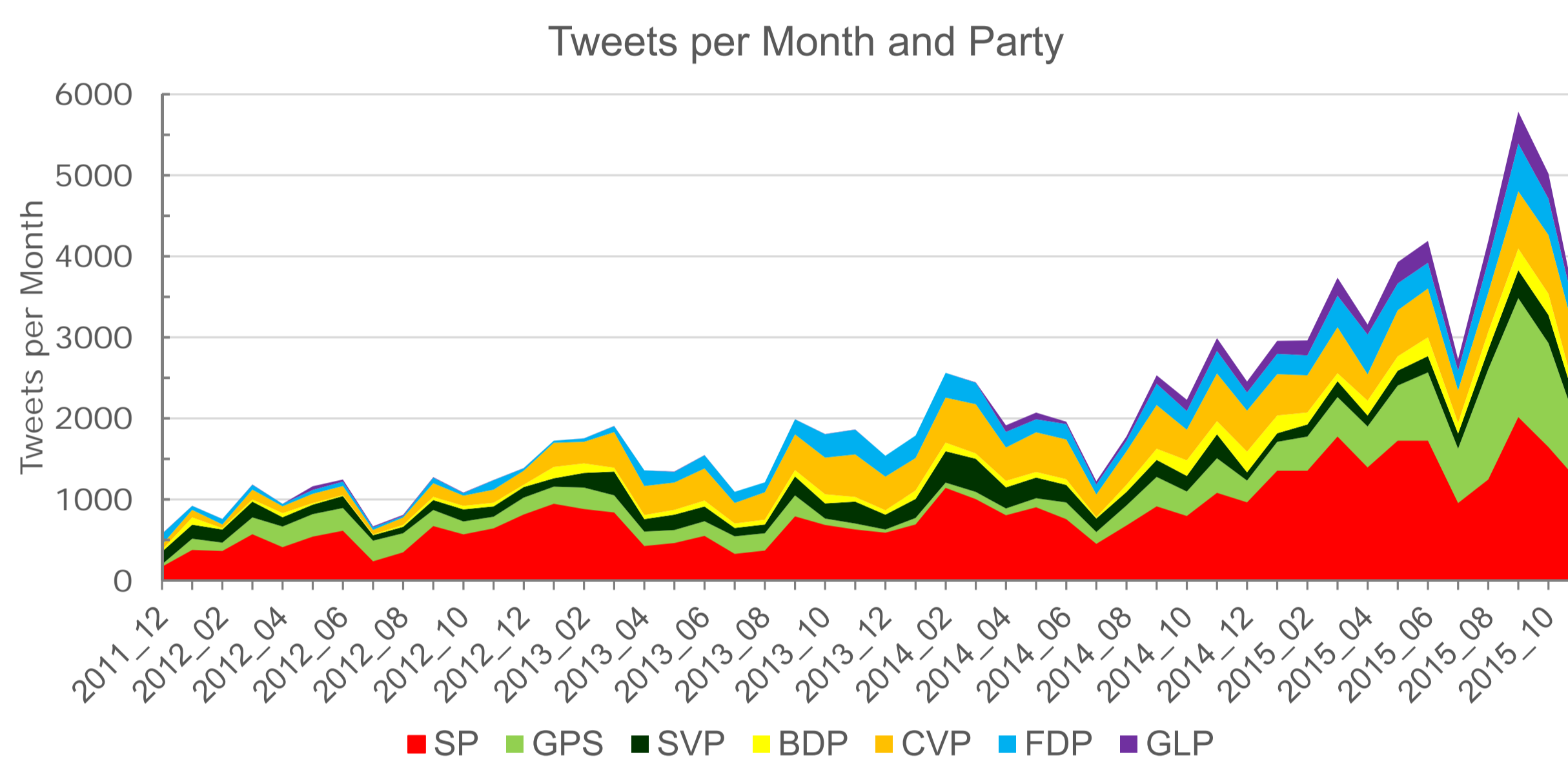


Multilanguage sentiment-analysis of Twitter data on the example of Swiss politicians

Twitter is increasingly important for political discussions

Out of the 246 members of the Swiss parliament a total of 119 have an Account on Twitter. Combined they write around **25'000 tweets per year** and most of them are about political topics. This makes these tweets a prime candidate for further text analysis.

However, people in Switzerland tweet in German, French, Italian, English and sometimes even Swiss German. This is a big problem for most current NLP algorithms.

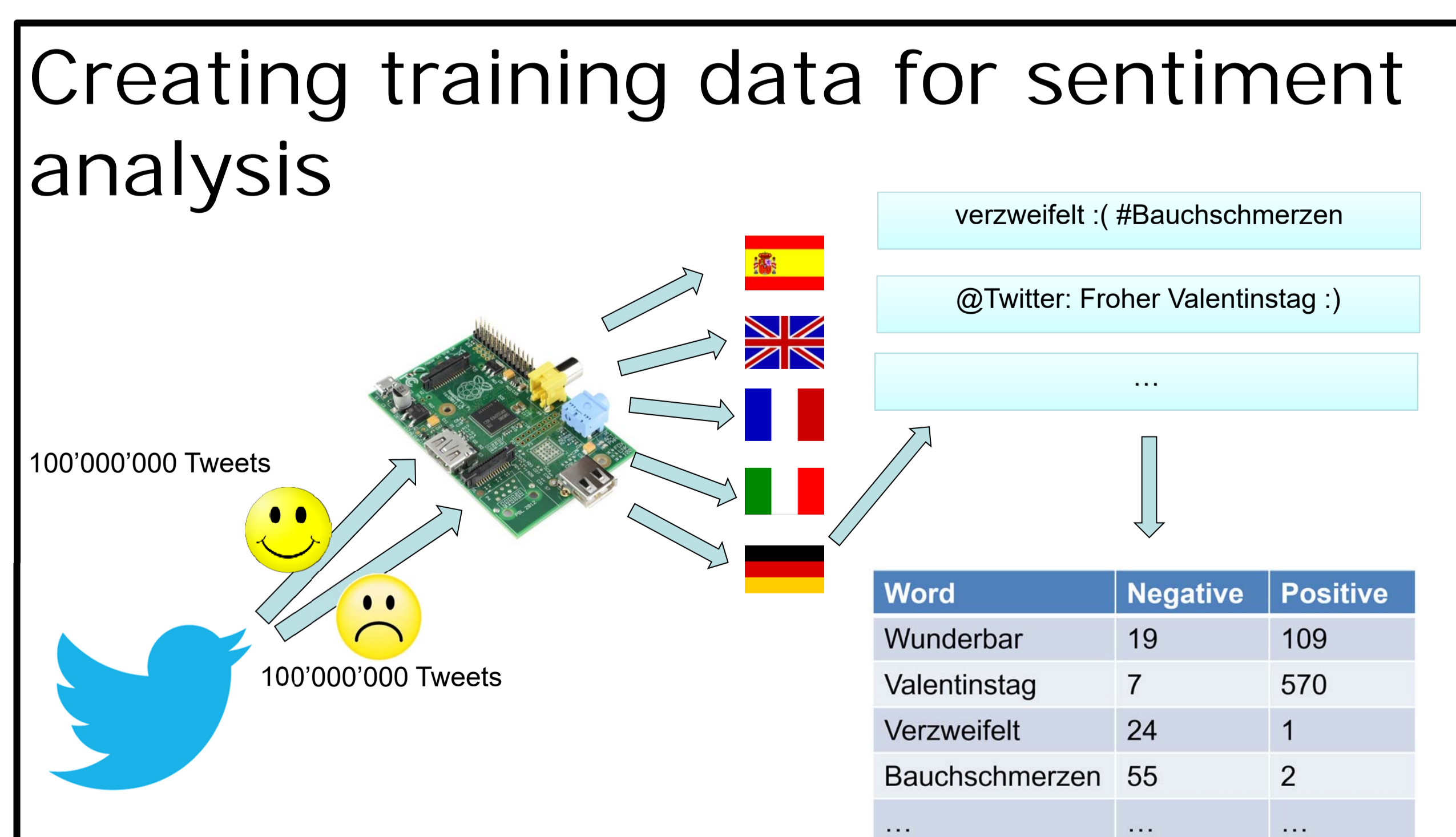


Pic. 1: The usage of Twitter has by members of the parliament has constantly risen during the 49th legislative period.

Emoticons as language independent indicators for sentiment

Of all the text messages sent on Twitter each day, approximately 5 - 10% contain an emoticon. These are used to express a wide range of different emotions and determine in most cases correctly the sentiment the author conveyed with his message, making it ideal to use as noisy labels for distant supervised learning.

Over a few months we have collected over 200 million tweets containing positive or negative emoticons. This corpus was then split by language and for each word the number of occurrences was counted, creating easily maintainable sentiment lexica for multiple languages.



Author: Lucas Brönnimann
Advisor: Prof. Dr. Manfred Vogel
University: University of Applied Sciences
Northwestern Switzerland

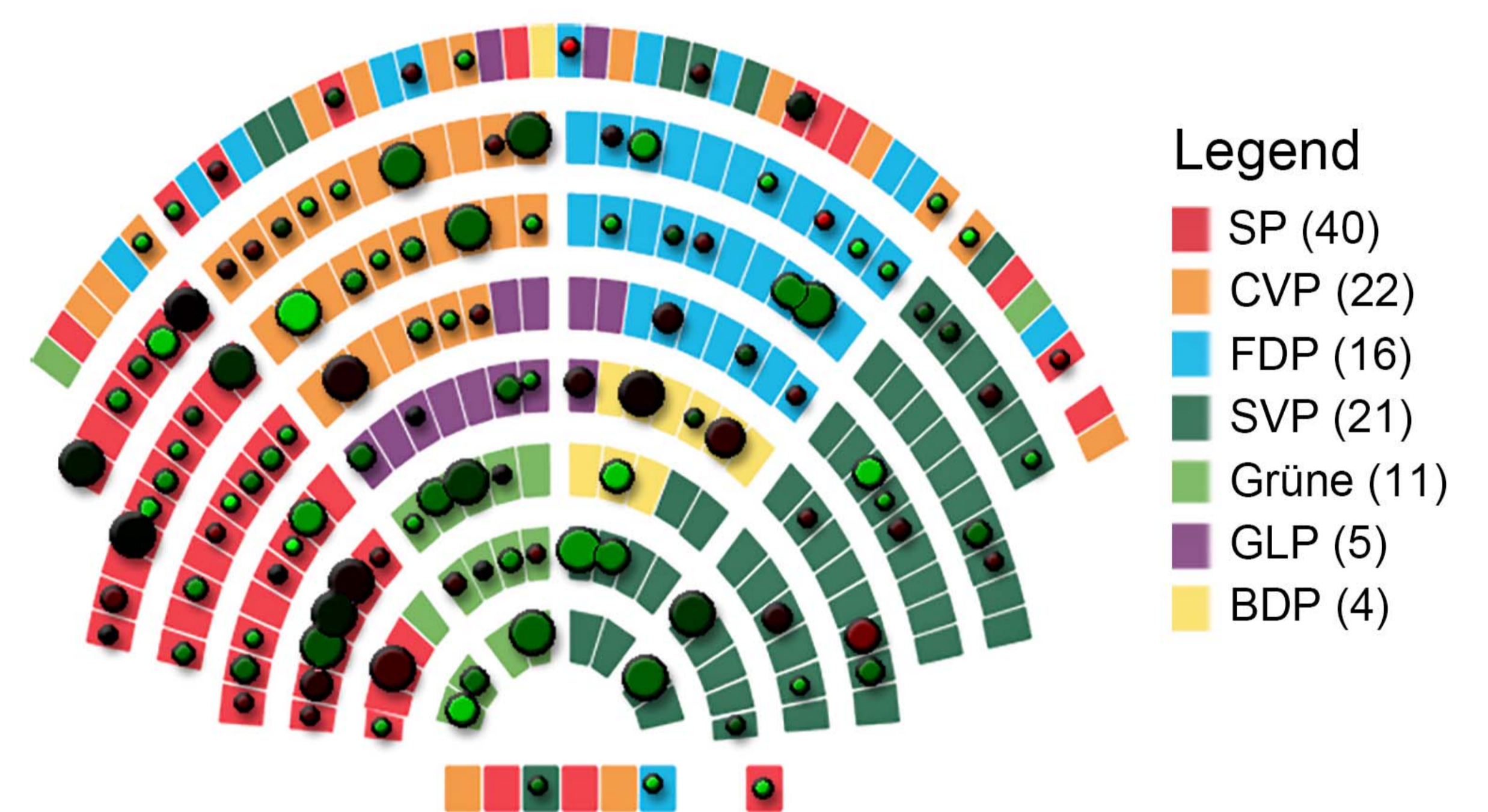
Word	Negative	Positive
rt	3'991'393	5'245'004
to	2'972'964	2'742'714
you	2'500'653	3'358'939
my	2'380'459	1'852'518
unusable	1'871	23
heartbreaking	6'619	158
prosperity	66	1'183

Table 1: Common words and their co-occurrences with positive and negative emoticons on Twitter in English.

Use the language specific sentiment lexicon to classify new tweets

To determine the sentiment of a new tweet (or other texts) various algorithms can be used based on the created sentiment lexicon. It has been shown that the quality of the classification is much more dependent on the training set than on the algorithm. Therefore a simple Naive Bayes approach was chosen to create a classifier which can categorize tweets into positive, neutral and negative.

The ~100'000 tweets of all members of the Swiss parliament in the 49th legislative period (2012-2015) provide a good use case for this classifier. Just by using the sentiment and typical topics the party membership can be correctly guessed in 54% of all Twitter users of the parliament.



Pic. 2: The size of the circle represents the Twitter activity of the user and the color indicates the average sentiment (green: positive, red: negative)



Pic. 3: Word Cloud based on the last 50 tweets by all members of the parliament. Green words are positive, red negative.

Conclusion

The strong correlation between usage of emoticons and the sentiment of a text is a helpful property for sentiment analysis. Applied on Swiss politicians a new view on the political landscape becomes possible, leading to surprising insights for newspapers and more transparency for voters.