Group 9
Refugees in media

Simon Schauß, Lukas Härtel, Alexander Schneider

Agenda

- Research Question
- Methodology
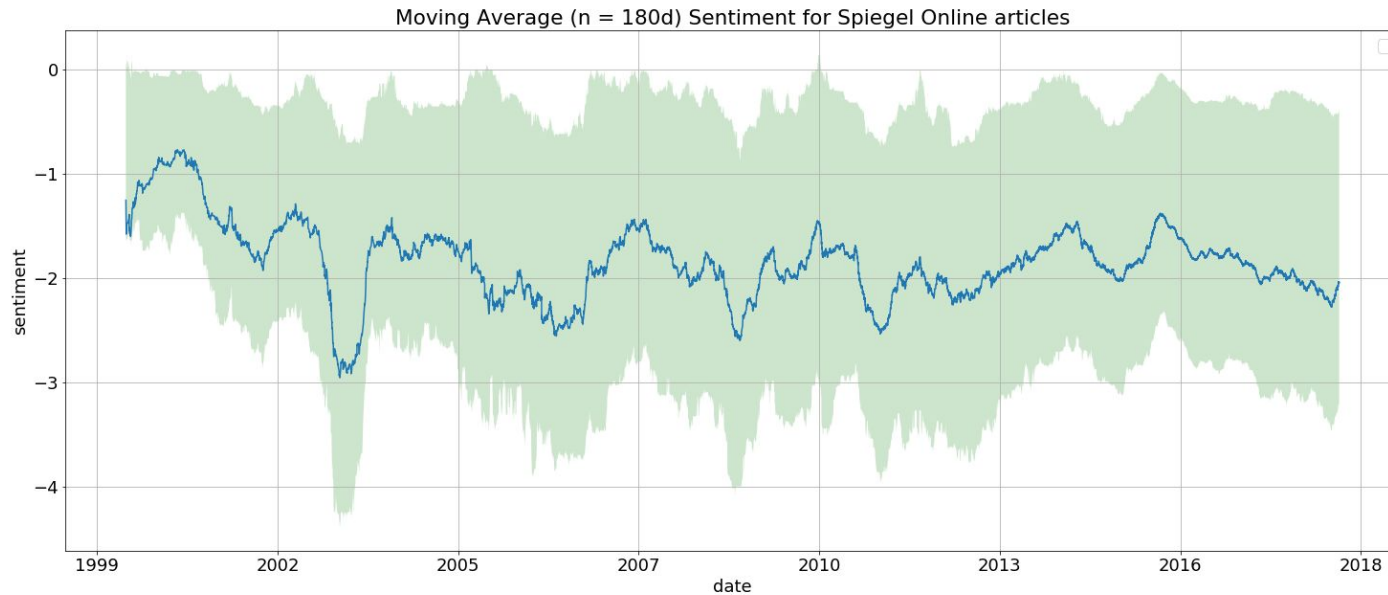- Results
- Limitations
- Related Work
- References

Research Question

- How has the general coverage on refugees developed over time
- How has the coverage on refugees regarding specific countries developed over time?
- What are the most common positive and negative sentiment-words used in this context?
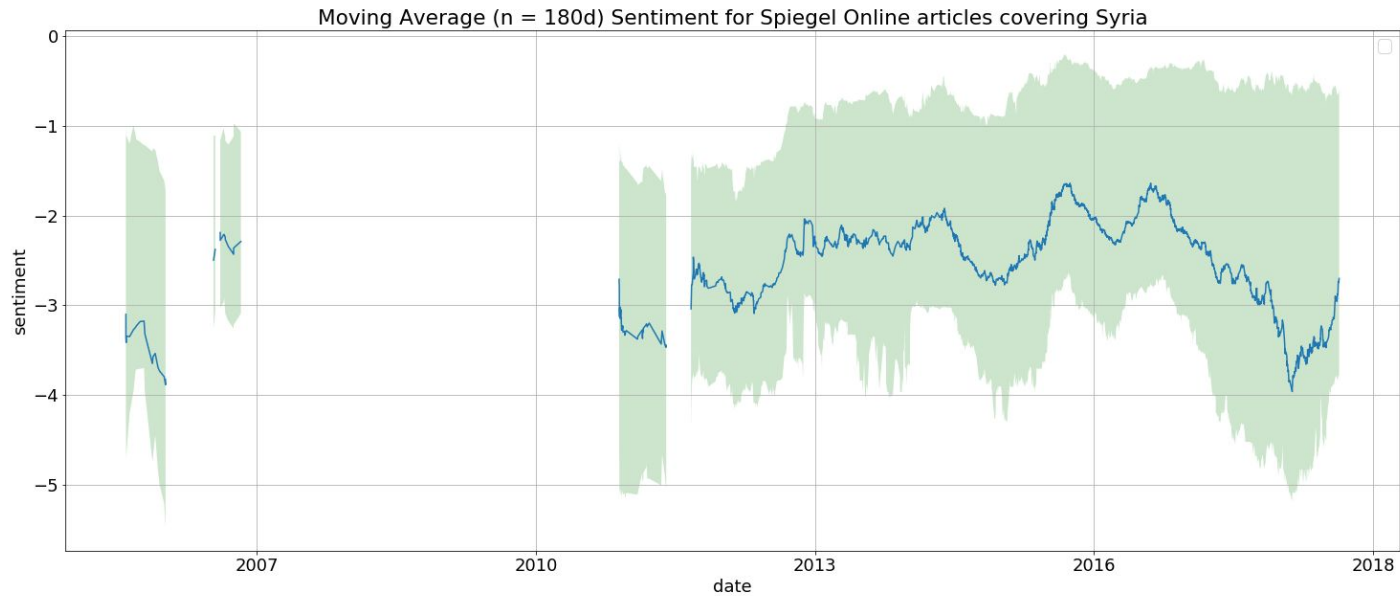
Methodology

- Download all articles with their publishing date from Spiegel Online (spiegel.de) since 01/01/2000
- Filter articles related to "Flüchtlinge" and synonyms ("Asylant", "Asylbewerber", "Asylsuchender", "Heimatvertriebener", "Migrant")
- Tokenize articles with trained tokenizer
- Tag sentences with STTS POS
- Evaluate sentiment polarity for articles based on SentiWS dictionary
- Find most common (sentiment)-words
- Plot moving average sentiment over time using a interactive line chart

Results - General coverage

Moving Average (n = 180d) Sentiment for Spiegel Online articles

Results - Coverage specific country - Syria

Moving Average (n = 180d) Sentiment for Spiegel Online articles covering Syria

Results - Sentiment-words - Spiegel

Rank	Word (pos)	english	total freq	Word (neg)	english	total freq
1	neue	new	10580	Flüchtlinge	refugee	28339
2	neuen	new (flex)	8721	Ende	end	10634
3	gut	good	8553	Flüchtlingen	refugee (flex)	7914
4	große	big, great	5442	Krieg	war	4952
5	großen	big, great (flex)	5199	Gewalt	violence	4668
6	Hilfe	help	4171	Kritik	criticism	4203
7	klar	clean	3972	Angst	fear	4146
8	schnell	fast	3717	Problem	problem, issue	4107
9	möglich	possible	3461	Flucht	escape	3797
10	Sicherheit	safety	3270	Kampf	fight	3617

20.430.815 words in total

Results - Sentiment-words - Spiegel

Rank	Word (pos)	english	rel. freq	Word (neg)	english	rel. freq
1	neue	new	0,0518%	Flüchtlinge	refugee	0,1390%
2	neuen	new (flex)	0,0427%	Ende	end	0,0520%
3	gut	good	0,0419%	Flüchtlingen	refugee (flex)	0,0387%
4	große	big, great	0,0266%	Krieg	war	0,0242%
5	großen	big, great (flex)	0,0254%	Gewalt	violence	0,0228%
6	Hilfe	help	0,0204%	Kritik	criticism	0,0206%
7	klar	clean	0,0194%	Angst	fear	0,0203%
8	schnell	fast	0,0182%	Problem	problem, issue	0,0201%
9	möglich	possible	0,0169%	Flucht	escape	0,0186%
10	Sicherheit	safety	0,0160%	Kampf	fight	0,0177%

20.430.815 words in total

Results - Sentiment-words - NYTimes

Rank	Word (pos)	total freq	Word (neg)	total freq
1	work	23554	war	31404
2	well	14682	killed	11477
3	peace	12447	illegal	10256
4	right	12104	conflict	9337
5	good	11999	attacks	8929
6	support	10951	death	7510
7	great	8458	attack	7156
8	better	8029	crisis	7052
9	best	7956	issue	7016
10	free	7380	died	6619

27.535.806 words in total

Results - Sentiment-words - NYTimes

Rank	Word (pos)	rel. freq	Word (neg)	rel. freq
1	work	0,0855%	war	0,1140%
2	well	0,0533%	killed	0,0417%
3	peace	0,0452%	illegal	0,0372%
4	right	0,0439%	conflict	0,0339%
5	good	0,0436%	attacks	0,0324%
6	support	0,0398%	death	0,0273%
7	great	0,0307%	attack	0,0260%
8	better	0,0292%	crisis	0,0256%
9	best	0,0289%	issue	0,0255%
10	free	0,0268%	died	0,0240%

27.535.806 words in total

Limitations

- NYTimes sentiment results for now neglected due to comparability and scope of Spiegel Online analysis (Only one newspaper considered)
- No set countries (We got all necessary data to build a interactive solution)
- Few articles for some cases to get representative results
- SentiWS has opinionated polarities

Related work

- SentiWS v1.8c resource for sentiment analysis (Universität Leipzig) http://wortschatz.uni-leipzig.de/de/download
- Opinion Lexicon by Bing Liu https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
- STTS POS (Universität Stuttgart) http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html
- Data Visualization https://west.uni-koblenz.de/sites/default/files/03_visualizations.pdf
- Sun (2003): Creating Meaningful Visualizations from Natural Language https://pdfs.semanticscholar.org/4d05/d7e9ff6cd461b6a95d5acc5349a8c8f510ee.pdf

References

- ipywidgets (plotting): https://ipywidgets.readthedocs.io/en/latest/
- nltk (language processing): https://www.nltk.org/
- BeautifulSoup (text scraping): https://www.crummy.com/software/BeautifulSoup/
- pyspark (parallelization): https://spark.apache.org/