



Representation of ▼ Russia in British press

before and after Annexation of Crimea

Natalia Kuzminykh

History



- The Crimean Peninsula, was annexed by the Russian Federation between February and March 2014;
- Ukraine and many other countries condemned the annexation and consider it to be a violation of international law;
- This event led to implementation of strict measures against the country (e.g. sanctions, suspension from G8 etc.)

Related works

- Nevinskaitė, L., 2017. Framing Russia as Threat in the Lithuanian Media: Corpus Analysis. *Lituanus*, 63(4).
 - Differences:
 - Topic: Russian foreign politics
 - Source: Lithuanian media
 - Period: from March 2014 (annexation of Crimea) to February 2015
 - Results: Russia as aggressor and potential threat to Lithuanian national security

Resources and tools:

- **Press:** BBC, The Telegraph
- **Tools:** Python 3.8., The Natural Language Toolkit (NLTK), Sketch Engine;
- **Parsing packages:** JusText, BeautifulSoup



Press

- The analysis was based on BBC and The Telegraph media sources;
 - Excluded: The Times and The Guardian
- Time period:
 - 1) From 20 Feb 2013 to 19 Feb 2014
 - 2) From 20 Feb 2014 to 20 Feb 2015
- Key words for search: “Russia”, “Moscow”;
 - Excluded: broadcastings, videos, radio, music, photo reports, sport results etc.

Building a corpus

- To automate the process of data collection, there was used a web scraping tool built with Python libraries:
 - Initial search: **Justext**
 - *Accuracy: 84.7% for BBC and 97.1% for Telegraph*
 - Second search: **BeautifulSoup**

Building a corpus

■ Process:

1. Creating a web scraper;
2. Extracting text from web pages to .txt files;
 - on this stage the corpus was separated on 4 parts:
Before (BBC/Telegraph) & After (BBC/Telegraph);
3. Combining data from BBC and Telegraph
 - Before and After corpora
 - **NB:** each .txt file has a name according to the theme and a source from where it was obtained: e.g.
BBC_WN_5.txt or *Tel_Sp_7.txt*
 - One can differentiate the following topics: *Sport; Business & Finance; World News; UK news; Blogs; Travels (only for Telegraph), and Other.*

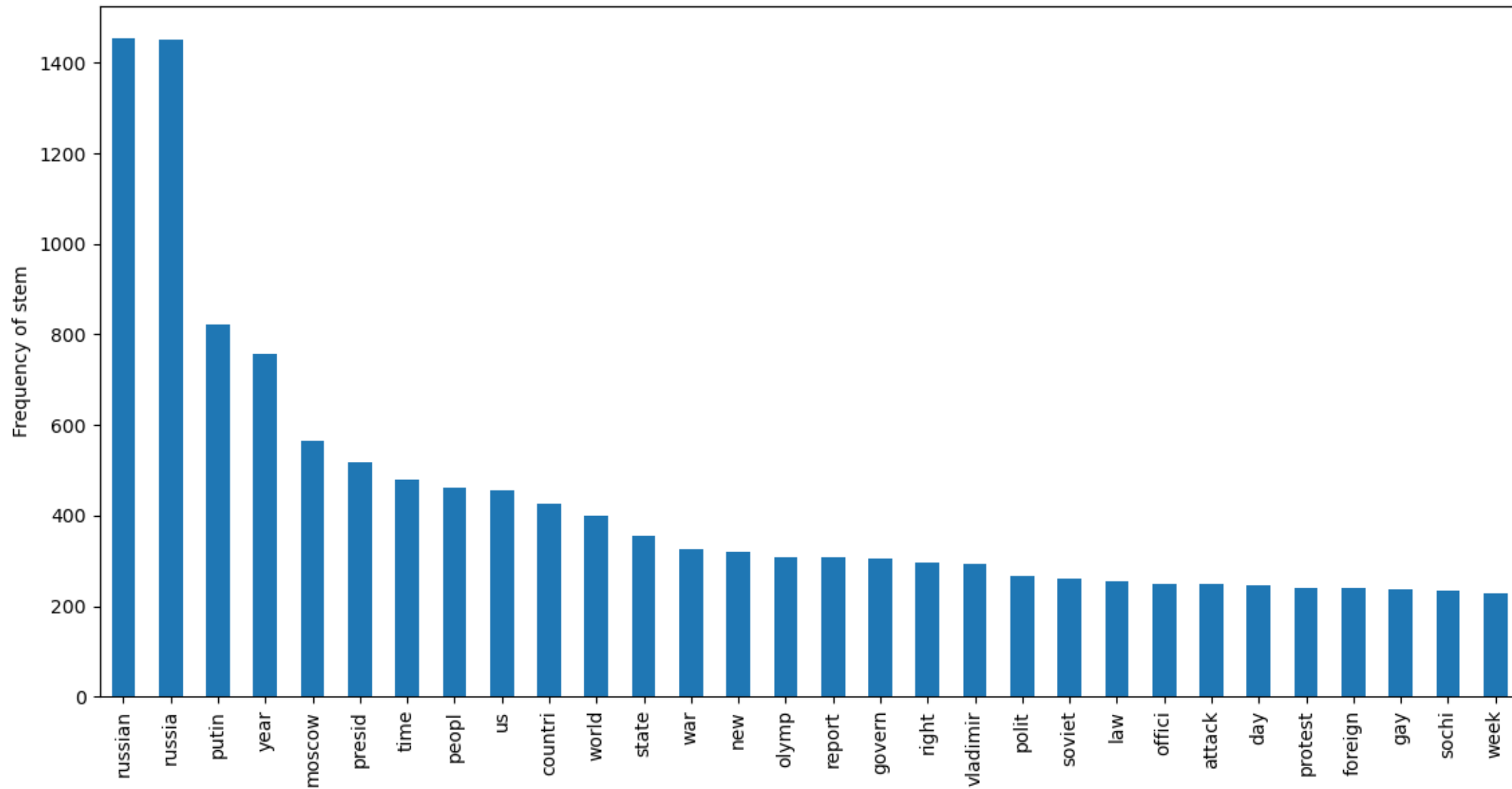
	BBC		The Telegraph	
Articles:	451		532	
	Before	After	Before	After
Articles:	183	268	264	268
Corpus				
	Before Annexation		After Annexation	
Articles:	447		536	
Tokens:	256,741		345,183	

Collected

Analysis

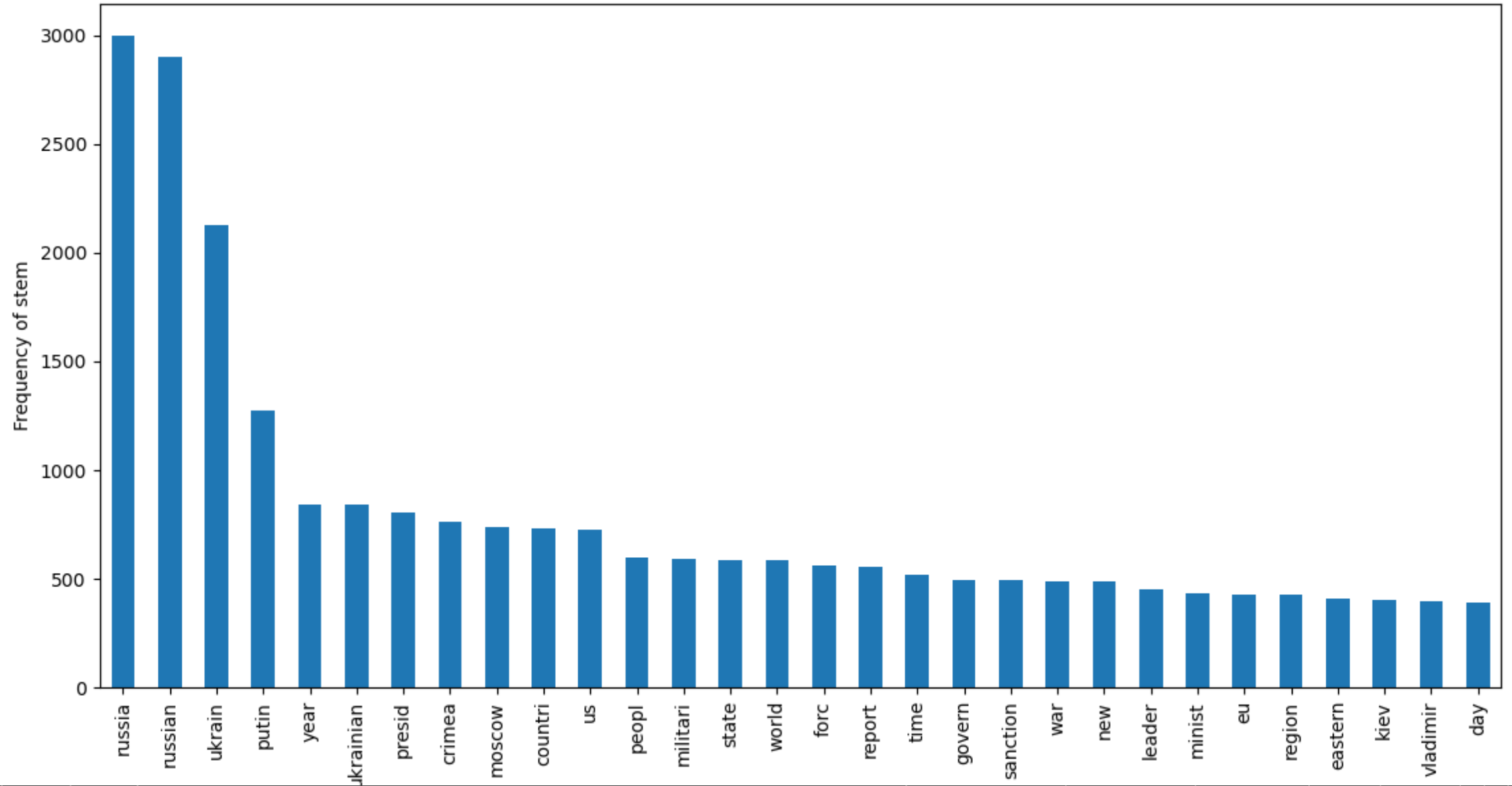
- The analysis was conducted with a help of NLTK and Sketch Engine tools;
 - **NLTK**: Frequency of use of words within the corpus after stemming and removal of stopwords;
 - **Sketch Engine**: Keyword extraction, Distributional Thesaurus, Collocation comparison etc.

Frequency of the most common stems



Frequency of the most common stems (Before Annexation)

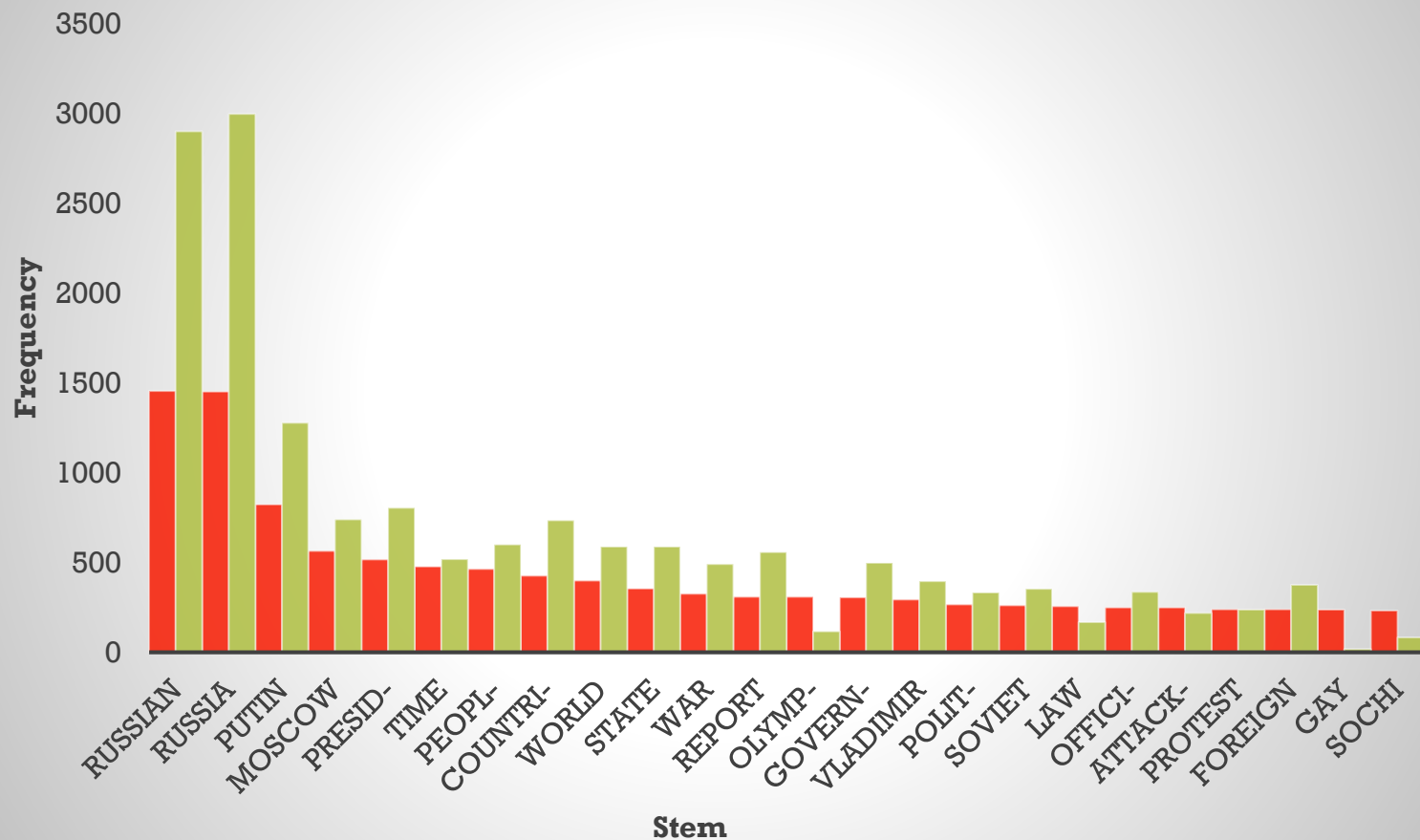
Frequency of the most common stems



Frequency of the most common stems (After Annexation)

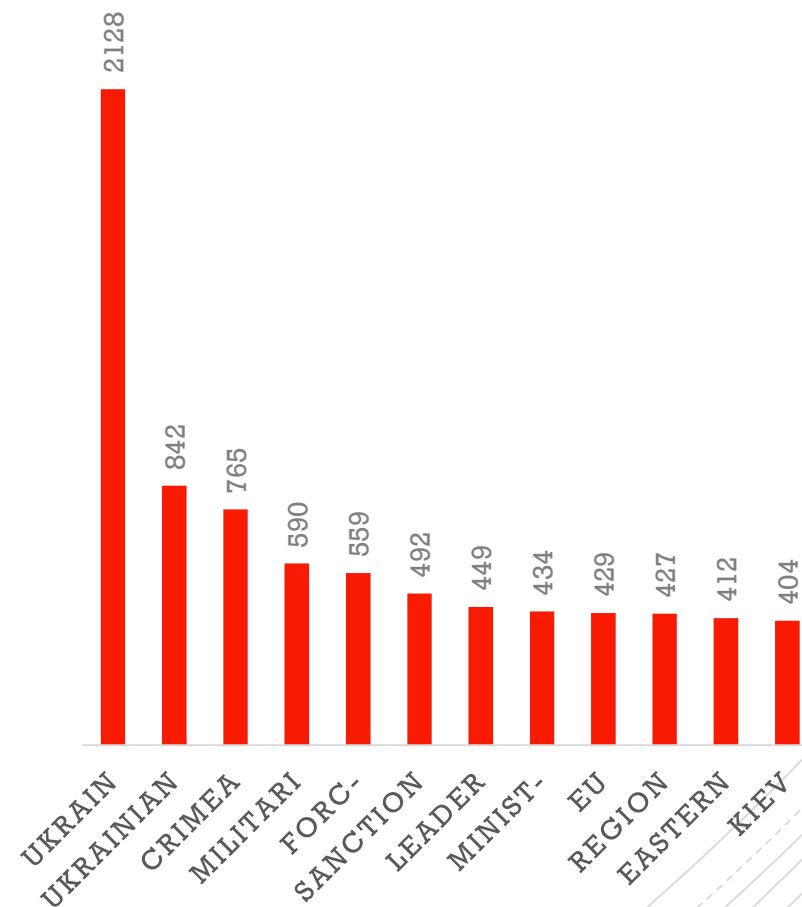
Comparison

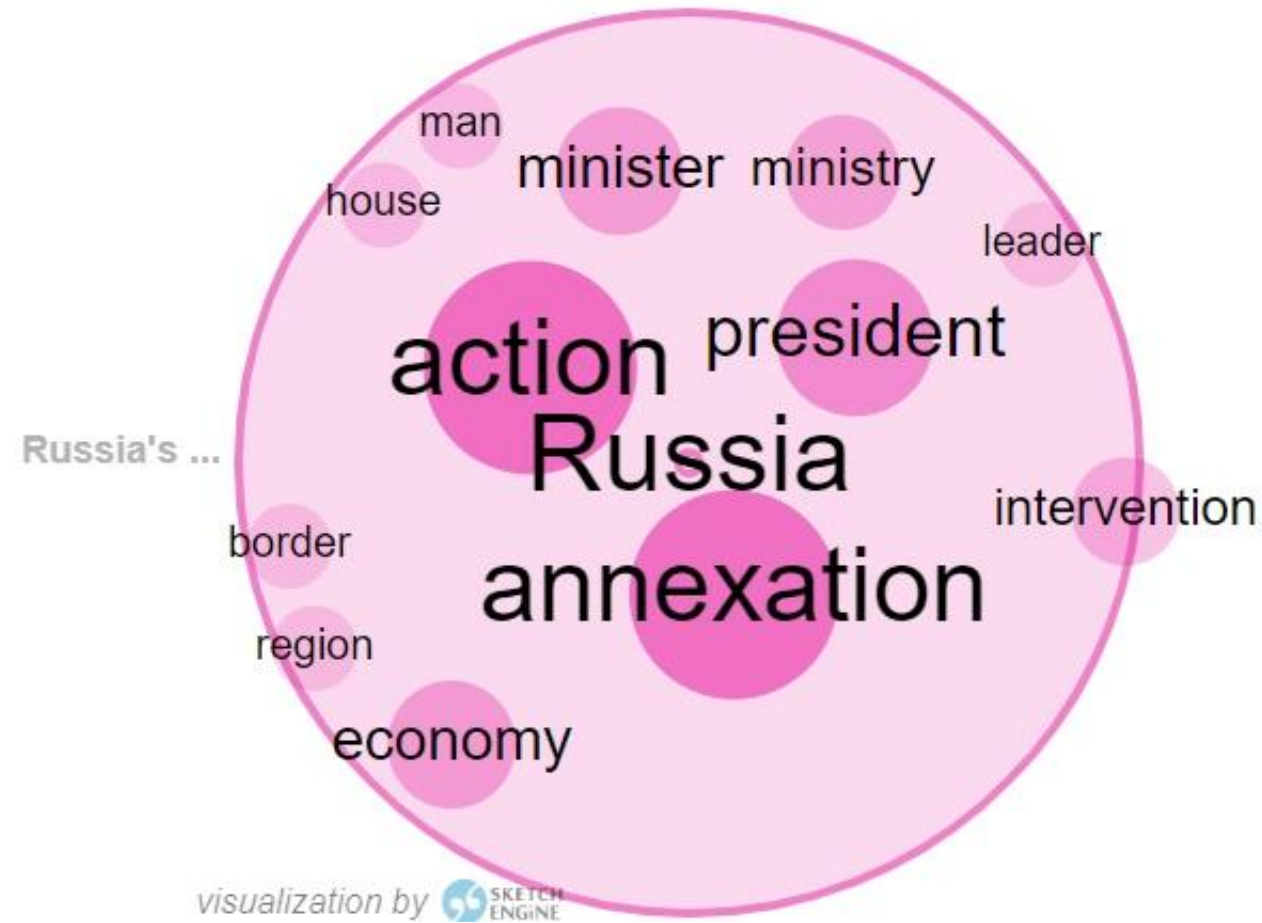
Frequency of the most common stems



UPTICK

■ After





Word collocations ('Russia')



Word collocations ('Russian')



Distributional Thesaurus

- an automatically generated list of synonyms or words belonging to the same category

Frequency?				
	Word	Focus	Reference	
1	Putin	804	76,926	...
2	Berezovsky	173	2,259	...
3	<u>Sochi</u>	232	19,898	...
4	Kremlin	191	25,782	...
5	Vladimir	294	59,897	...
6	<u>Magnitsky</u>	88	2,215	...
7	Khodorkovsky	102	7,369	...
8	Yanukovich	79	7,313	...
9	Browder	61	2,396	...
10	Moscow	553	219,973	...

Frequency?				
	Word	Focus	Reference	
11	<u>Snowden</u>	143	44,858	...
12	Sergei	88	22,316	...
13	Russia	1,440	721,285	...
14	Tolokonnikova	46	1,356	...
15	oligarch	62	10,352	...
16	Filin	43	713	...
17	Dmitry	68	15,776	...
18	Bolshoi	44	4,104	...
19	Russian	1,259	783,044	...
20	BST	51	10,655	...

Keyword extraction (Before Annexation)

- Reference corpus: English Web
- single word units which are typical for a corpus/ext or which define its content or topic

Frequency?					Frequency?				
Word	Focus	Reference			Word	Focus	Reference		
1 <u>Crimea</u>	763	7,376	...		11 <u>separatist</u>	239	22,598	...	
2 Putin	1,269	76,926	...		12 <u>Crimean</u>	148	6,488	...	
3 <u>Ukraine</u>	2,121	169,327	...		13 <u>MH17</u>	110	0	...	
4 Nato	289	8,397	...		14 Vladimir	395	59,897	...	
5 <u>Donetsk</u>	240	4,913	...		15 <u>Poroshenko</u>	97	161	...	
6 <u>Ukrainian</u>	762	86,849	...		16 Russia	2,963	721,285	...	
7 <u>Kiev</u>	392	34,546	...		17 <u>Luhansk</u>	89	280	...	
8 Kremlin	293	25,782	...		18 <u>rouble</u>	111	6,274	...	
9 Yanukovich	176	7,313	...		19 noframe	77	119	...	
10 Litvinenko	145	2,106	...		20 <u>Sevastopol</u>	82	2,747	...	

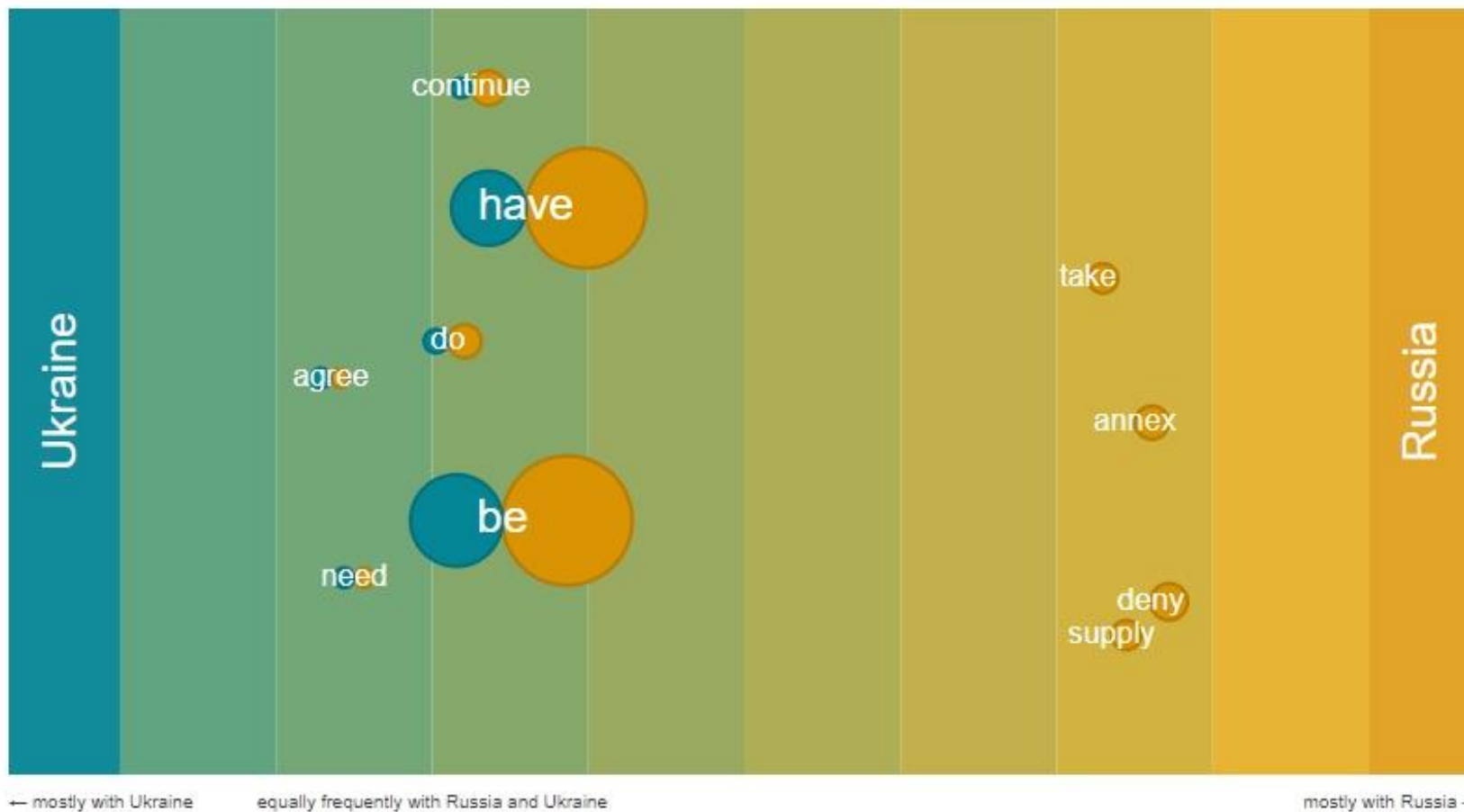
Keyword extraction (After Annexation)

- Reference corpus: English Web
- single word units which are typical for a corpus/ext or which define its content or topic



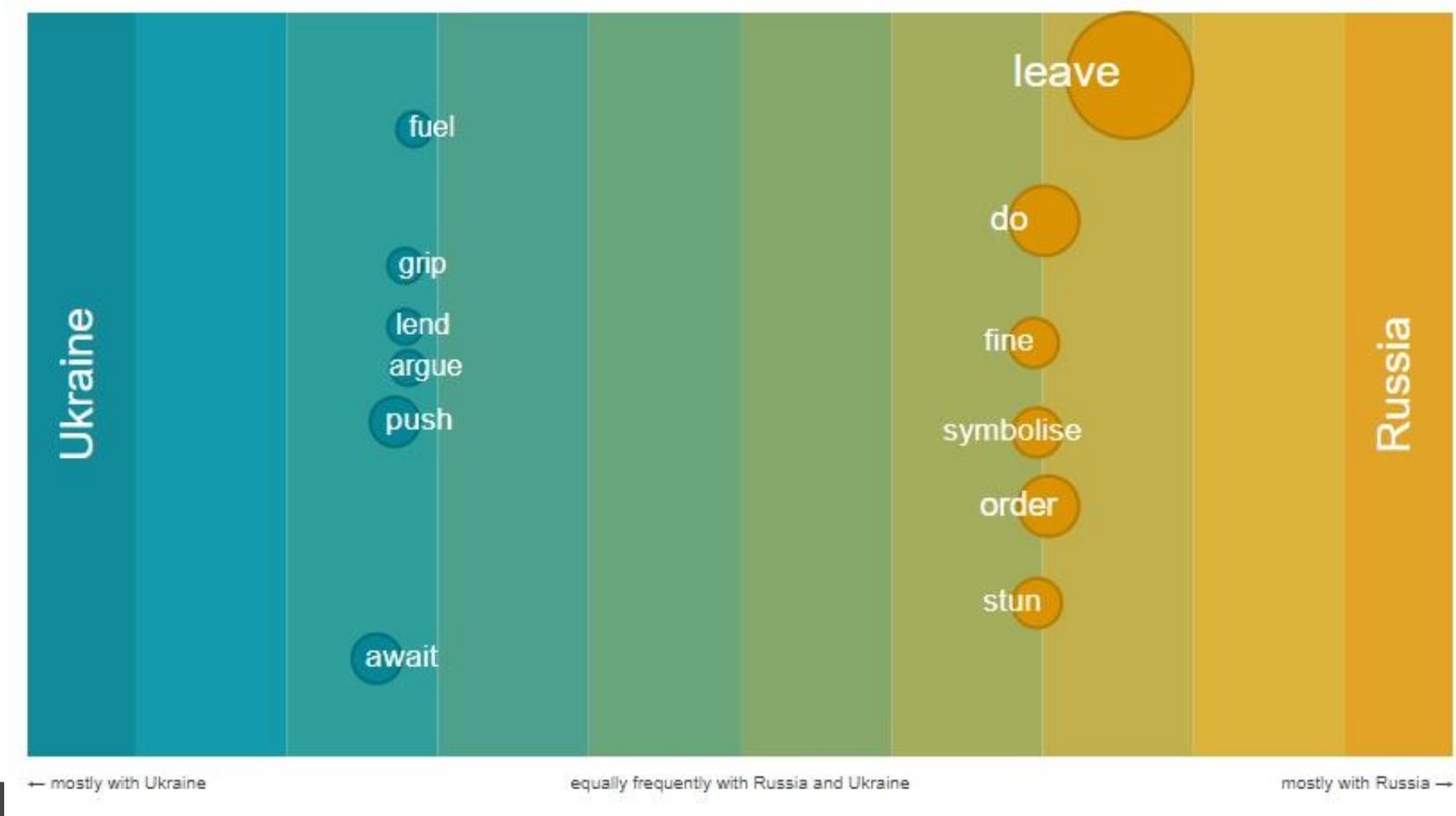
Collocations of two words (Before Annexation)

- *Verb collocations with 'Russia/Ukraine' as subject*



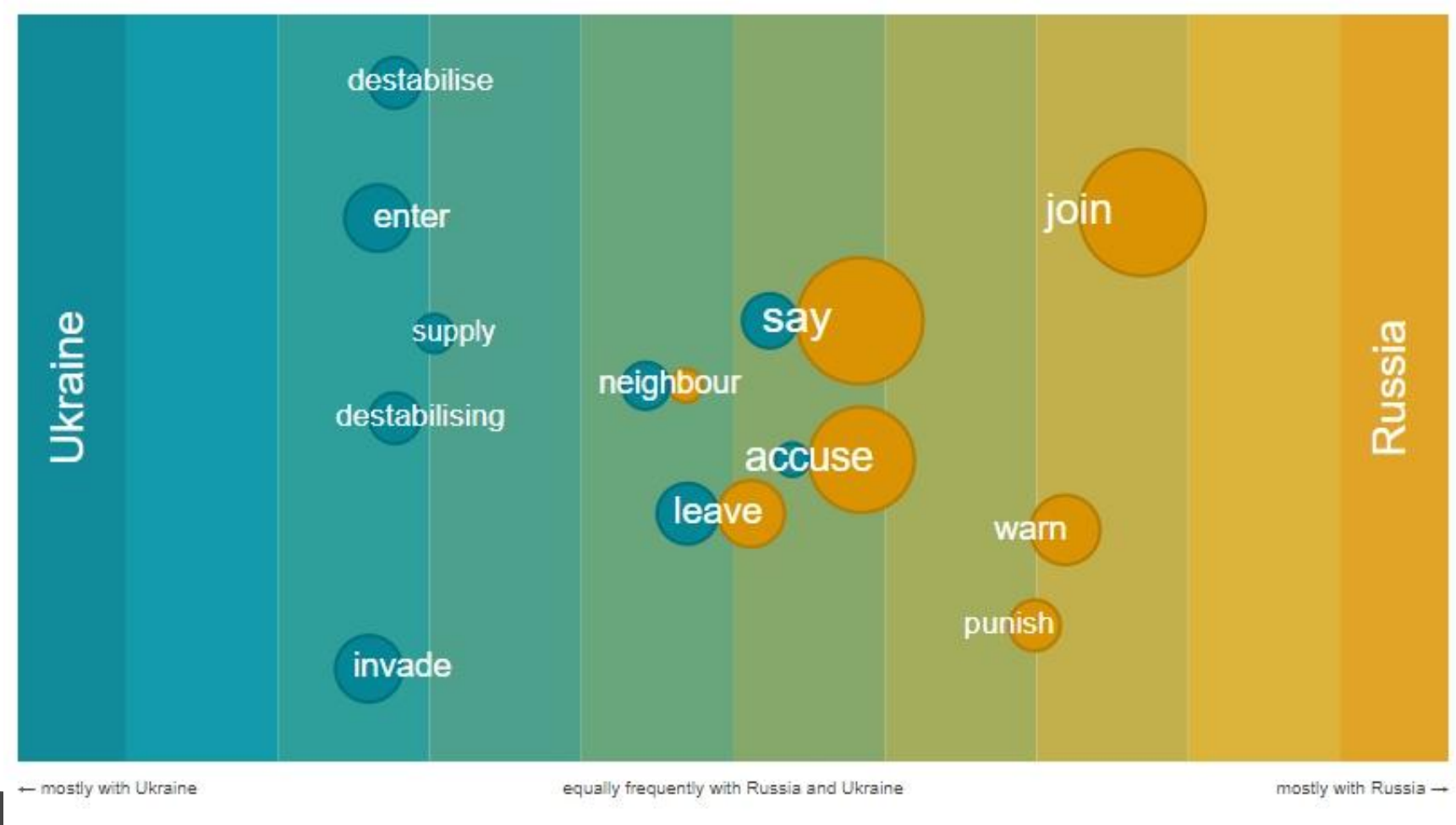
Collocations of two words (After Annexation)

- *Verb collocations with 'Russia/Ukraine' as subject*



Collocations of two words (Before Annexation)

- *Verb collocations with 'Russia/Ukraine' as object*



Collocations of two words (After Annexation)

- *Verb collocations with 'Russia/Ukraine' as object*

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a sense of motion or a stylized globe. On the left side, there is a large red speech bubble with a tail pointing downwards.

Further research

- Sentiment analysis of the corpus;
- Extending the chosen period;
- Extending corpus by including posts from blogs
e.g. twitter

The background features a series of concentric circles in a light red color. A solid red arrow points downwards from the left side, intersecting one of the circles. The text is centered in a bold, red, sans-serif font.

THANK YOU FOR
YOUR ATTENTION