NLP

Find your favorite news source and grab the article text.

1. Show the most common words in the article.
2. Show the most common words under a part of speech. (i.e. NOUN: {'Bob':12, 'Alice':4,})
3. Find a subject/object relationship through the dependency parser in any sentence.
4. Show the most common Entities and their types.
5. Find Entites and their dependency (hint: entity.root.head)
6. Find the most similar words in the article

Note: Yes, the notebook from the video is not provided, I leave it to you to make your own :) it's your final assignment for the semester. Enjoy!

Install Packages

```
In [1]:  import spacy
         processor = spacy.load('en_core_web_sm')
```

Source: Foreign Affairs Article Link: https://www.foreignaffairs.com/russian-federation/vladimir-putin-age-chaos

```
In [2]:  # Importing Article Text
         text = """'"After Russian President Vladimir Putin launched an invasion of Ukraine in

         The following months have only ratcheted up the pressure. The Ukrainian front has prov

         Prigozhin's uprising captured the world's attention—and deeply disturbed Moscow's elit

         Stay informed.
         In-depth analysis delivered weekly.
         These events were unprecedented in contemporary Russia. And yet they do not seem to ha

         Don't be fooled: this ostensible resilience to bad news and the seeming indifference t

         Prigozhin's mutiny has pushed the situation to an extreme and may pave the way for the

         FRAGILE STATE
         In the months leading up to Prigozhin's rebellion, Russia found itself in the unexpect

         The Kremlin's propensity to downplay seemingly shocking events is in line with how Put

         Prigozhin's growing discontent preceding the mutiny failed to alarm these insiders. Ev


         Prigozhin leaving Rostov-on-Don, Russia, June 2023
         Alexander Ermochenko / Reuters
         These events served as a sobering revelation: Putin had misjudged Prigozhin and his ou
```

In the wake of the rebellion, it has become much harder for the Kremlin to project an

The Kremlin's mishandling of the war, compounded by Prigozhin's ensuing mutiny, has ma

SOMETHING ROTTEN
Many observers are used to viewing Putin's regime as the product of a social contract

But Putin's passivity in the face of internal military threats and his aloof stance ma

Anti-elite sentiment also propelled the rise of Prigozhin. He had been gaining visibil

Ordinary Russians were struck by Prigozhin's brazen confrontation with the Defense Mir

It is becoming harder for the Kremlin to sweep bad news under the carpet.
Many Western observers have suggested that these military troubles will push the elite

If anything, the country is becoming more committed to the fight—not in pursuit of imp

The mutiny and the events that preceded it suggest that the regime may be much less re

Putin, lulled into complacency by his conviction that people love him and the elites a

THE HARDENING LINE
This situation plays squarely into the hands of Russia's hard-liners, a camp that cons

Since the invasion, the Russian state has marginalized antiwar forces and left no room

Worse, the old guard now has little to say about wartime realities, with great uncerta

As their star wanes, a new generation of hawks is rising. Some of these new hawks are

Checking debris after a reported Ukrainian drone attack in Moscow, August 2023
Evgenia Novozhenina
Some senior members of the establishment have benefited from the hawkish turn, however

The clash of hawks, old and new, will shape Russia's response to its struggles in Ukra

In truth, Putin and those old ideologues close to him, such as Patrushev, are becoming

GOING OFF SCRIPT
The system has started to learn to operate independently of Putin. This development do

Other factions are already moving into the space opened by Putin's weakness. Putin cou

Contrary to what analysts might have anticipated would follow Prigozhin's uprising—att

At the same time, however, the political class is shifting its attention inward to add

The war has left Russia grasping for certainties in an exceedingly uncertain world.
Ordinary Russians still seem to support the war and back Putin, but they are also beco

For the foreseeable future, the Kremlin will be wrestling simultaneously with divergin

If previously, domestic affairs were secondary to the dominant military agenda, the re

In [3]: article_text = processor(text)

1. Show the most common words in the article.

```python
In [4]:  # Importing Additional Libraries
         from collections import Counter
         import pandas as pd
         import numpy as np
         import string
```

```python
In [5]:  # Define stop words, punctuation
         stop_words = set(processor.Defaults.stop_words)
         punctuation = set(string.punctuation)
```

```python
In [6]:  # Create a counting function (with exceptions for stop words, punctuation)
         counts = Counter(token.text for token in article_text if not token.is_stop and not tok
```

```python
In [7]:  # Convert Dictionary to Dataframe
         word_counts = pd.DataFrame.from_dict(counts, orient='index', columns=['Frequency'])
         word_counts = word_counts.sort_values(by='Frequency', ascending=False)
         word_counts.reset_index(inplace=True)
         word_counts.rename(columns={'index': 'Word'}, inplace=True)
```

```python
In [8]:  # Show Word Counts
         word_counts.iloc[0:20]
```

Out[8]:

	Word	Frequency
0	Putin	54
1	war	46
2	Prigozhin	37
3	Russian	33
4	Russia	27
5	military	25
6	Kremlin	25
7	Russians	23
8	state	21
9	Ukraine	20
10	—	19
11	regime	18
12	political	17
13	"	15
14	"	14
15	mutiny	14
16	elites	14
17	Wagner	13
18	Moscow	12
19	Ukrainian	12

The most common words are 'Putin', 'war', 'Prigozhin', 'Russian'/'Russia' and 'military'.

1. Show the most common words under a part of speech. (i.e. NOUN: {'Bob':12, 'Alice':4,})

In [9]:
```
# Create Counter For Parts of Speech
parts_of_speech_counts = Counter((token.text, token.pos_) for token in article_text if
```

In [10]:
```
part_of_speech_tags = {token.pos_ for token in article_text}
part_of_speech_tags
```

Out[10]:   {'ADJ',
             'ADP',
             'ADV',
             'AUX',
             'CCONJ',
             'DET',
             'NOUN',
             'NUM',
             'PART',
             'PRON',
             'PROPN',
             'PUNCT',
             'SCONJ',
             'SPACE',
             'SYM',
             'VERB',
             'X'}

In [11]:
```python
# For each part of speech, count occurances of a token if that token matched the giver
for part in part_of_speech_tags:
    filtered_text = [token.text for token in article_text if token.pos_ == part]
    counts_for_part_of_speech = Counter(filtered_text)
    print(part, counts_for_part_of_speech)
    print() # I am adding extra space for Readability
```

ADV Counter({'more': 34, 'even': 14, 'now': 12, 'also': 11, 'less': 9, 'only': 7, 'most': 6, 'as': 6, 'increasingly': 5, 'far': 5, 'Even': 4, 'however': 4, 'simply': 4, 'instead': 4, 'well': 4, 'much': 4, 'Then': 3, 'yet': 3, 'Instead': 3, 'so': 3, 'inward': 3, 'still': 3, 'already': 3, 'never': 3, 'long': 3, 'rather': 3, 'just': 2, 'likely': 2, 'further': 2, 'gradually': 2, 'Now': 2, 'overtly': 2, 'entirely': 2, 'completely': 2, 'harder': 2, 'ever': 2, 'previously': 2, 'conspicuously': 2, 'quickly': 2, 'exceedingly': 2, 'really': 1, 'surprisingly': 1, 'actually': 1, 'prior': 1, 'ultimately': 1, 'deeply': 1, 'publicly': 1, 'ruthlessly': 1, 'freely': 1, 'weekly': 1, 'openly': 1, 'Indeed': 1, 'necessarily': 1, 'somewhat': 1, 'home': 1, 'briefly': 1, 'startlingly': 1, 'barely': 1, 'seemingly': 1, 'enough': 1, 'Consequently': 1, 'often': 1, 'privately': 1, 'consistently': 1, 'After': 1, 'all': 1, 'Just': 1, 'emotionally': 1, 'spontaneously': 1, 'rarely': 1, 'perhaps': 1, 'altogether': 1, 'resolutely': 1, 'desperately': 1, 'Unfortunately': 1, 'ahead': 1, 'best': 1, 'at': 1, 'worst': 1, 'away': 1, 'certainly': 1, 'Together': 1, 'squarely': 1, 'radically': 1, 'virtually': 1, 'everywhere': 1, 'physically': 1, 'intellectually': 1, 'regardless': 1, 'thereby': 1, 'closely': 1, 'on': 1, 'endlessly': 1, 'notably': 1, 'easily': 1, 'directly': 1, 'More': 1, 'especially': 1, 'internally': 1, 'Regardless': 1, 'inevitably': 1, 'effectively': 1, 'seriously': 1, 'independently': 1, 'before': 1, 'harshly': 1, 'eventually': 1, 'actively': 1, 'eagerly': 1, 'ago': 1, 'mainly': 1, 'typically': 1, 'primarily': 1, 'widely': 1, 'generally': 1, 'simultaneously': 1})

PART Counter({'to': 95, ''s': 84, 'not': 22, 'To': 2, ''': 2, 'n't': 1})

NUM Counter({'one': 5, '2022': 4, '2023': 2, '23': 1, 'two': 1, '2014': 1, 'eight': 1, '19': 1, '68': 1, '48': 1, 'five': 1, '45': 1, '21': 1, 'four': 1})

SPACE Counter({'\n\n': 31, '\n': 12, '\n\n\n': 2})

PUNCT Counter({',': 296, '.': 181, '-': 31, '—': 19, '"': 15, '"': 14, ';': 12, ':': 11, '(': 10, ')': 10, "'": 1, '"': 1, '[': 1, ']': 1})

ADP Counter({'of': 164, 'in': 98, 'to': 64, 'for': 44, 'with': 42, 'as': 33, 'on': 32, 'by': 21, 'from': 14, 'about': 14, 'at': 13, 'into': 12, 'In': 12, 'up': 8, 'against': 8, 'among': 8, 'out': 7, 'toward': 6, 'down': 6, 'than': 6, 'after': 6, 'under': 4, 'At': 4, 'between': 3, 'over': 3, 'across': 2, 'before': 2, 'during': 2, 'By': 2, 'like': 2, 'onto': 1, 'On': 1, 'within': 1, 'With': 1, 'inside': 1, 'As': 1, 'Unlike': 1, 'unlike': 1, 'OFF': 1, 'For': 1})

CCONJ Counter({'and': 173, 'but': 12, 'or': 8, 'But': 5, 'And': 4, 'both': 2, 'either': 1})

X Counter({'de': 1, 'facto': 1})

SCONJ Counter({'that': 34, 'as': 8, 'if': 6, 'how': 5, 'for': 4, 'why': 4, 'Since': 3, 'when': 3, 'since': 3, 'If': 2, 'than': 2, 'After': 1, 'although': 1, 'Despite': 1, 'With': 1, 'until': 1, 'Although': 1, 'despite': 1, 'where': 1, 'after': 1, 'because': 1})

VERB Counter({'become': 13, 'becoming': 10, 'growing': 8, 'left': 5, 'try': 5, 'lead': 4, 'have': 4, 'come': 4, 'including': 4, 'deal': 4, 'make': 4, 'exposed': 3, 'revealed': 3, 'supposed': 3, 'following': 3, 'opened': 3, 'allowed': 3, 'remains': 3, 'saw': 3, 'remain': 3, 'maintain': 3, 'made': 3, 'began': 3, 'appear': 3, 'want': 3, 'According': 3, 'perceived': 3, 'appears': 3, 'launched': 2, 'seemed': 2, 'carry': 2, 'getting': 2, 'caused': 2, 'move': 2, 'seem': 2, 'continue': 2, 'daring': 2, 'sweep': 2, 'begun': 2, 'ignore': 2, 'turning': 2, 'leading': 2, 'found': 2, 'used': 2, 'hit': 2, 'seized': 2, 'sought': 2, 'endure': 2, 'convinced': 2, 'believe': 2, 'challenge': 2, 'advancing': 2, 'shot': 2, 'underestimating': 2, 'posed': 2, 'escalating': 2, 'helped': 2, 'grown': 2, 'seeking': 2, 'given': 2, 'demonstrate': 2, 'inspired': 2, 'support': 2, 'show': 2, 'say': 2, 'increased': 2, 'beginning': 2, 'suggested': 2, 'hoped': 2, 'gaining': 2, 'trusted': 2, 'dared': 2, 'explains': 2, 'reported': 2, 'starte

d': 2, 'see': 2, 'led': 2, 'detached': 2, 'leave': 2, 'seek': 2, 'win': 2, 'settle': 2, 'use': 2, 'find': 2, 'avoid': 2, 'occur': 2, 'go': 2, 'known': 2, 'shifting': 2, 'address': 2, 'grasping': 2, 'acted': 1, 'changed': 1, 'foundered': 1, 'tightened': 1, 'Irked': 1, 'engulfed': 1, 'produced': 1, 'rebound': 1, 'ratcheted': 1, 'provided': 1, 'connected': 1, 'attacked': 1, 'raided': 1, 'carried': 1, 'seizing': 1, 'sending': 1, 'racing': 1, 'shooting': 1, 'killing': 1, 'captured': 1, 'disturbed': 1, 'brokered': 1, 'struggle': 1, 'understand': 1, 'condemned': 1, 'hosted': 1, 'Stay': 1, 'delivered': 1, 'ruffled': 1, 'happened': 1, 'complain': 1, 'date': 1, 'reshuffled': 1, 'arrested': 1, 'fooled': 1, 'change': 1, 'empowering': 1, 'eclipsing': 1, 'view': 1, 'underestimate': 1, 'addressing': 1, 'refuse': 1, 'linked': 1, 'pushed': 1, 'pave': 1, 'radicalized': 1, 'turn': 1, 'bring': 1, 'transforming': 1, 'open': 1, 'built': 1, 'contend': 1, 'watching': 1, 'aligned': 1, 'crossed': 1, 'neighboring': 1, 'enduring': 1, 'focused': 1, 'advanced': 1, 'commented': 1, 'preferring': 1, 'delegate': 1, 'downplay': 1, 'maintained': 1, 'seated': 1, 'brim': 1, 'achieves': 1, 'taking': 1, 'panic': 1, 'speaking': 1, 'bragged': 1, 'offering': 1, 'let': 1, 'cited': 1, 'preceding': 1, 'failed': 1, 'alarm': 1, 'initiated': 1, 'continued': 1, 'happening': 1, 'achieving': 1, 'channeling': 1, 'supervising': 1, 'leaving': 1, 'served': 1, 'misjudged': 1, 'turned': 1, 'intervene': 1, 'spark': 1, 'embittered': 1, 'inflamed': 1, 'fighting': 1, 'endorsed': 1, 'project': 1, 'hugging': 1, 'surprised': 1, 'seen': 1, 'acting': 1, 'held': 1, 'shook': 1, 'striking': 1, 'wanted': 1, 'interpreted': 1, 'experience': 1, 'compounded': 1, 'ensuing': 1, 'thwart': 1, 'stoked': 1, 'ensure': 1, 'mind': 1, 'stop': 1, 'infiltrating': 1, 'reveals': 1, 'stab': 1, 'viewing': 1, 'guarantees': 1, 'granting': 1, 'manage': 1, 'eclipsed': 1, 'accompanied': 1, 'delegated': 1, 'threatens': 1, 'causes': 1, 'owing': 1, 'admire': 1, 'reelected': 1, 'ruling': 1, 'are': 1, 'doubt': 1, 'fulfill': 1, 'targeted': 1, 'live': 1, 'associated': 1, 'serve': 1, 'wake': 1, 'compelling': 1, 'helping': 1, 'responding': 1, 'propelled': 1, 'operated': 1, 'elevated': 1, 'ranked': 1, 'struck': 1, 'described': 1, 'going': 1, 'gave': 1, 'took': 1, 'missing': 1, 'noted': 1, 'filled': 1, 'disappeared': 1, 'barricaded': 1, 'push': 1, 'crave': 1, 'tend': 1, 'waging': 1, 'quashing': 1, 'understood': 1, 'fueled': 1, 'increase': 1, 'conclude': 1, 'drawn': 1, 'said': 1, 'believed': 1, 'last': 1, 'adapting': 1, 'steeling': 1, 'halt': 1, 'subdued': 1, 'suppressed': 1, 'advocating': 1, 'according': 1, 'interpreting': 1, 'backed': 1, 'conducted': 1, 'struggling': 1, 'frightened': 1, 'preceded': 1, 'suggest': 1, 'frazzled': 1, 'failing': 1, 'perplexed': 1, 'trembling': 1, 'fly': 1, 'crumbles': 1, 'trying': 1, 'investigate': 1, 'shocked': 1, 'patch': 1, 'lulled': 1, 'love': 1, 'do': 1, 'arrest': 1, 'clamp': 1, 'complicating': 1, 'dislodging': 1, 'plays': 1, 'consists': 1, 'advocate': 1, 'tightening': 1, 'hunting': 1, 'placing': 1, 'footing': 1, 'accumulate': 1, 'reconcile': 1, 'overt': 1, 'end': 1, 'supporting': 1, 'aiding': 1, 'conduct': 1, 'feel': 1, 'perform': 1, 'marginalized': 1, 'cracking': 1, 'begin': 1, 'enacting': 1, 'outlawing': 1, 'gain': 1, 'bolder': 1, 'supplant': 1, 'foster': 1, 'promote': 1, 'contributed': 1, 'precipitated': 1, 'stripped': 1, 'has': 1, 'looming': 1, 'promoted': 1, 'removed': 1, 'rising': 1, 'announced': 1, 'annexing': 1, 'tasked': 1, 'overseeing': 1, 'destroyed': 1, 'spend': 1, 'occupied': 1, 'demonstrating': 1, 'include': 1, 'involved': 1, 'observing': 1, 'drone': 1, 'seize': 1, 'spout': 1, 'Checking': 1, 'benefited': 1, 'beef': 1, 'access': 1, 'alter': 1, 'put': 1, 'write': 1, 'give': 1, 'raging': 1, 'deploy': 1, 'shape': 1, 'challenges': 1, 'faces': 1, 'evolve': 1, 'wait': 1, 'hostage': 1, 'called': 1, 'secure': 1, 'boost': 1, 'avert': 1, 'organize': 1, 'contribute': 1, 'regarding': 1, 'hawkish': 1, 'saving': 1, 'presume': 1, 'fail': 1, 'wage': 1, 'incomprehensible': 1, 'calling': 1, 'establish': 1, 'managed': 1, 'advocated': 1, 'argued': 1, 'using': 1, 'bolster': 1, 'considering': 1, 'discussing': 1, 'looks': 1, 'committed': 1, 'bears': 1, 'GOING': 1, 'learn': 1, 'operate': 1, 'reflect': 1, 'emerging': 1, 'reflects': 1, 'neglected': 1, 'aging': 1, 'falter': 1, 'enabled': 1, 'dropping': 1, 'demanding': 1, 'allowing': 1, 'dares': 1, 'attack': 1, 'moving': 1, 'learning': 1, 'pandering': 1, 'limiting': 1, 'knows': 1, 'feeding': 1, 'longing': 1, 'depict': 1, 'waiting': 1, 'upend': 1, 'existing': 1, 'incurring': 1, 'received': 1, 'honing': 1, 'shaping': 1, 'directing': 1, 'encouraging': 1, 'benefits': 1, 'Maintaining': 1, 'help': 1, 'achieve': 1, 'lacks': 1, 'agreeing': 1, 'focus': 1, 'outmaneuvering': 1, 'anticipated': 1, 'follow': 1, 'consolidate': 1, 'dismantle': 1, 'integrate': 1, 'proposed': 1, 'formalizing': 1, 'irritate': 1, 'discussed': 1, 'separating': 1, 'seeks': 1, 'acquire': 1, 'coupled': 1, 'ignite': 1, 'localized': 1, 'expected': 1, 'dismantled': 1,

'seems': 1, 'allow': 1, 'vying': 1, 'focusing': 1, 'becomes': 1, 'highlight': 1, 'blame': 1, 'making': 1, 'conducting': 1, 'frustrated': 1, 'showing': 1, 'feeling': 1, 'enjoy': 1, 'mask': 1, 'back': 1, 'pose': 1, 'wrestling': 1, 'diverging': 1, 'deepening': 1, 'intensifying': 1, 'marked': 1, 'shrinking': 1, 'play': 1, 'reacts': 1})

SYM Counter({'/': 1})

ADJ Counter({'Russian': 33, 'military': 20, 'political': 17, 'many': 14, 'Ukrainian': 12, '-': 12, 'new': 11, 'such': 11, 'domestic': 10, 'social': 8, 'Western': 8, 'anti': 8, 'internal': 7, 'own': 6, 'ordinary': 6, 'significant': 6, 'more': 6, 'private': 5, 'mercenary': 5, 'high': 5, 'public': 5, 'little': 4, 'other': 4, 'old': 4, 'hawkish': 4, 'senior': 4, 'close': 4, 'last': 3, 'much': 3, 'top': 3, 'ongoing': 3, 'hard': 3, 'loyal': 3, 'Many': 3, 'geopolitical': 3, 'powerful': 3, 'pro': 3, 'recent': 3, 'same': 3, 'antiwar': 3, 'conservative': 3, 'pragmatic': 3, 'strategic': 3, 'usual': 2, 'short': 2, 'key': 2, 'shocking': 2, 'open': 2, 'bad': 2, 'unwelcome': 2, 'deep': 2, 'long': 2, 'ruthless': 2, 'volatile': 2, 'unpredictable': 2, 'Other': 2, 'passive': 2, 'resilient': 2, 'presidential': 2, 'special': 2, 'major': 2, 'large': 2, 'aloof': 2, 'general': 2, 'ready': 2, 'hostile': 2, 'nationalist': 2, 'very': 2, 'influential': 2, 'current': 2, 'elite': 2, 'fat': 2, 'provincial': 2, 'brutal': 2, 'anxious': 2, 'desperate': 2, 'chaotic': 2, 'war': 2, 'national': 2, 'traditional': 2, 'former': 2, 'deputy': 2, 'entire': 2, 'enormous': 2, 'personal': 2, 'true': 2, 'armed': 2, 'uncertain': 2, 'complicated': 1, 'successful': 1, 'tremendous': 1, 'flexible': 1, 'intolerable': 1, 'tangible': 1, 'good': 1, 'Unknown': 1, 'swift': 1, 'Belarusian': 1, 'informed': 1, 'unprecedented': 1, 'contemporary': 1, 'sure': 1, 'overall': 1, 'stable': 1, 'ostensible': 1, 'seeming': 1, 'deceptive': 1, 'difficult': 1, 'profound': 1, 'underway': 1, 'ultranationalist': 1, 'habitual': 1, 'immediate': 1, 'cohesive': 1, 'lacking': 1, 'outright': 1, 'disorderly': 1, 'dangerous': 1, 'unexpected': 1, 'early': 1, 'annual': 1, 'unknown': 1, 'paramilitary': 1, 'various': 1, 'continual': 1, 'mute': 1, 'rare': 1, 'controllable': 1, 'cool': 1, 'fine': 1, 'harmful': 1, 'prudent': 1, 'cautious': 1, 'strong': 1, 'late': 1, 'useful': 1, 'certain': 1, 'clear': 1, 'sobering': 1, 'voluble': 1, 'detached': 1, 'harder': 1, 'unflappable': 1, 'unscheduled': 1, 'unprepared': 1, 'adoring': 1, 'acute': 1, 'irresponsible': 1, 'weak': 1, 'potential': 1, 'deeper': 1, 'stringent': 1, 'official': 1, 'regional': 1, 'near': 1, 'ambivalent': 1, 'famous': 1, 'wealthy': 1, 'sorry': 1, 'rich': 1, 'just': 1, 'affluent': 1, 'parasitic': 1, 'Kremlin': 1, 'involved': 1, 'Anti': 1, 'independent': 1, 'Bakhmut': 1, 'important': 1, 'first': 1, 'stratospheric': 1, 'spoiled': 1, 'local': 1, 'simple': 1, 'warm': 1, 'deserted': 1, 'broader': 1, 'challenging': 1, 'determined': 1, 'inefficient': 1, 'instructive': 1, 'highest': 1, 'tough': 1, 'committed': 1, 'imperial': 1, 'resolute': 1, 'decisive': 1, 'consistent': 1, 'latest': 1, 'slight': 1, 'frustrated': 1, 'lethargic': 1, 'unimaginable': 1, 'repressive': 1, 'necessary': 1, 'tolerant': 1, 'minor': 1, 'suspicious': 1, 'indulgent': 1, 'privileged': 1, 'luxurious': 1, 'distant': 1, 'pressured': 1, 'liberal': 1, 'minded': 1, 'massive': 1, 'patriotic': 1, 'larger': 1, 'active': 1, 'hardcore': 1, 'younger': 1, 'older': 1, 'antiliberal': 1, 'spiritual': 1, 'Worse': 1, 'wartime': 1, 'great': 1, 'absolute': 1, 'decent': 1, 'dismal': 1, 'confrontational': 1, 'jingoistic': 1, 'nitty': 1, 'young': 1, 'chief': 1, 'prime': 1, 'fetched': 1, 'bizarre': 1, 'catastrophic': 1, 'main': 1, 'suppressed': 1, 'administrative': 1, 'hundredth': 1, 'evil': 1, 'real': 1, 'physical': 1, 'responsible': 1, 'darker': 1, 'fractious': 1, 'Most': 1, 'tentative': 1, 'indiscriminate': 1, 'rancorous': 1, 'competitive': 1, 'obsolete': 1, 'obsessed': 1, 'accurate': 1, 'radical': 1, 'alternative': 1, 'conceivable': 1, 'financial': 1, 'economic': 1, 'diplomatic': 1, 'Putin': 1, 'managerial': 1, 'genuine': 1, 'utter': 1, 'largest': 1, 'dynamic': 1, 'adept': 1, 'flattering': 1, 'innumerable': 1, 'international': 1, 'few': 1, 'extreme': 1, 'concurrent': 1, 'harsher': 1, 'firm': 1, 'capable': 1, 'narrow': 1, 'Contrary': 1, 'coordinated': 1, 'coherent': 1, 'exact': 1, 'martial': 1, 'wary': 1, 'autonomous': 1, 'militarized': 1, 'heavy': 1, 'additional': 1, 'interior': 1, 'possible': 1, 'further': 1, 'rival': 1, 'historic': 1, 'liberating': 1, 'popular': 1, 'effective': 1, 'back': 1, 'vulnerable': 1, 'clumsy': 1, 'unchanneled': 1, 'True': 1, 'foreseeable': 1, 'ineffective': 1, 'secondary': 1, 'dominant': 1, 'urgent': 1, 'bleak': 1, 'greater': 1, 'ideological': 1, 'stricter': 1, 'prominent': 1, 'predictable': 1, 'contradictory': 1, 'historical': 1})

```
AUX Counter({'has': 32, 'is': 25, 'are': 22, 'may': 19, 'will': 19, 'have': 17, 'wa
s': 16, 'had': 14, 'be': 14, 'were': 12, 'would': 9, 'could': 8, 'been': 7, 'can': 7,
'do': 2, 'does': 2, 'Do': 1, 'did': 1, 'might': 1})

DET Counter({'the': 380, 'a': 83, 'The': 29, 'an': 20, 'these': 10, 'A': 7, 'These':
6, 'another': 5, 'any': 4, 'this': 4, 'all': 4, 'no': 3, 'that': 3, 'those': 3, 'Thi
s': 3, 'Some': 2, 'That': 2, 'some': 2, 'every': 1, 'be—"the': 1, 'THE': 1, 'All': 1,
'No': 1})

PRON Counter({'that': 28, 'their': 21, 'its': 15, 'his': 15, 'who': 14, 'they': 12,
'it': 12, 'They': 9, 'he': 7, 'him': 7, 'It': 6, 'which': 6, 'them': 5, 'itself': 3,
'He': 3, 'themselves': 3, 'us': 3, 'His': 3, 'what': 3, 'anything': 2, 'nothing': 2,
'The': 2, 'something': 2, 'That': 2, 'those': 1, 'everything': 1, 'all': 1, 'himsel
f': 1, 'SOMETHING': 1, 'There': 1, 'someone': 1, 'she': 1, 'whom': 1, 'This': 1, 'bot
h': 1, 'Some': 1, 'the': 1})

NOUN Counter({'war': 44, 'state': 21, 'regime': 18, 'elites': 14, 'mutiny': 14, 'forc
es': 11, 'country': 10, 'events': 10, 'hawks': 10, 'security': 9, 'government': 9, 'o
fficials': 9, 'rebellion': 8, 'president': 8, 'authorities': 8, 'city': 7, 'people':
7, 'home': 6, 'public': 6, 'control': 6, 'sentiment': 6, 'invasion': 5, 'front': 5,
'months': 5, 'territory': 5, 'news': 5, 'uprising': 5, 'world': 5, 'elite': 5, 'end':
5, 'army': 5, 'military': 5, 'ideologues': 5, 'attacks': 5, 'percent': 5, 'future':
5, 'time': 5, 'figures': 5, 'society': 5, 'power': 4, 'region': 4, 'event': 4, 'respo
nse': 4, 'services': 4, 'situation': 4, 'developments': 4, 'challenges': 4, 'groups':
4, 'support': 4, 'conflict': 4, 'defense': 4, 'minister': 4, 'resources': 4, 'drone':
4, 'attack': 4, 'polling': 4, 'anxiety': 3, 'leader': 3, 'attention': 3, 'perceptio
n': 3, 'responsibility': 3, 'space': 3, 'dissent': 3, 'order': 3, 'patriotism': 3, 'a
dministration': 3, 'stability': 3, 'capacity': 3, 'sources': 3, 'command': 3, 'soldie
rs': 3, 'staff': 3, 'revolt': 3, 'force': 3, 'profile': 3, 'activities': 3, 'wake':
3, 'hands': 3, 'observers': 3, 'need': 3, 'media': 3, 'ranks': 3, 'year': 3, 'realit
y': 3, 'decline': 3, 'head': 3, 'course': 3, 'notion': 3, 'leadership': 3, 'sanction
s': 2, 'economy': 2, 'vulnerability': 2, 'sense': 2, 'pressure': 2, 'seizure': 2, 'dr
ones': 2, 'border': 2, 'company': 2, 'column': 2, 'troops': 2, 'number': 2, 'pilots':
2, 'part': 2, 'crisis': 2, 'hand': 2, 'carpet': 2, 'life': 2, 'liners': 2, 'guard':
2, 'realities': 2, 'weakening': 2, 'flaws': 2, 'risks': 2, 'term': 2, 'way': 2, 'weak
ness': 2, 'conflicts': 2, 'position': 2, 'regions': 2, 'line': 2, 'goals': 2, 'inside
rs': 2, 'operation': 2, 'approval': 2, 'ratings': 2, 'discontent': 2, 'concern': 2,
'reports': 2, 'commander': 2, 'product': 2, 'inaction': 2, 'stance': 2, 'chief': 2,
'failures': 2, 'outfit': 2, 'behavior': 2, 'members': 2, 'years': 2, 'citizens': 2,
'existence': 2, 'percentage': 2, 'institutions': 2, 'threats': 2, 'class': 2, 'blogge
rs': 2, 'visibility': 2, 'firm': 2, 'fighters': 2, 'politicians': 2, 'eyewitness': 2,
'cats': 2, 'powers': 2, 'action': 2, 'peace': 2, 'threat': 2, 'ambitions': 2, 'poll
s': 2, 'confidence': 2, 'words': 2, 'pundits': 2, 'opposition': 2, 'protests': 2, 'gr
ound': 2, 'generation': 2, 'ideas': 2, 'interests': 2, 'uncertainty': 2, 'lack': 2,
'policy': 2, 'affairs': 2, 'establishment': 2, 'turn': 2, 'opponents': 2, 'setbacks':
2, 'backdrop': 2, 'radicalization': 2, 'factions': 2, 'priorities': 2, 'fragmentatio
n': 2, 'problems': 2, 'certainties': 2, 'actions': 2, 'campaign': 1, 'autumn': 1, 'th
ings': 1, 'counterattack': 1, 'positions': 1, 'mobilization': 1, 'period': 1, 'strik
e': 1, 'bridge': 1, 'link': 1, 'mainland': 1, 'smoke': 1, 'flames': 1, 'redlines': 1,
'exception': 1, 'meantime': 1, 'assailants': 1, 'Paramilitaries': 1, 'aircraft': 1,
'dozen': 1, 'process': 1, 'resolution': 1, 'deal': 1, 'handling': 1, 'traitor': 1, 'n
egotiations': 1, 'depth': 1, 'analysis': 1, 'status': 1, 'quo': 1, 'lives': 1, 'gener
als': 1, 'brass': 1, 'personnel': 1, 'resilience': 1, 'indifference': 1, 'shifts': 1,
'attitude': 1, 'militarization': 1, 'touch': 1, 'inclination': 1, 'favor': 1, 'incide
nts': 1, 'extreme': 1, 'emergence': 1, 'Threats': 1, 'revelations': 1, 'entity': 1,
'rife': 1, 'contradictions': 1, 'predictability': 1, 'debate': 1, 'days': 1, 'parad
e': 1, 'attackers': 1, 'targets': 1, 'mayhem': 1, 'settlements': 1, 'shelling': 1, 'b
utton': 1, 'Television': 1, 'talk': 1, 'shows': 1, 'efficacy': 1, 'air': 1, 'defense
s': 1, 'narrative': 1, 'ruthlessness': 1, 'masters': 1, 'exceptions': 1, 'soil': 1,
```

```
'propensity': 1, 'belief': 1, 'path': 1, 'victory': 1, 'cues': 1, 'demeanor': 1, 'ave
rsion': 1, 'impact': 1, 'insider': 1, 'caveat': 1, 'objectives': 1, 'frustration': 1,
'ultranationalists': 1, 'addition': 1, 'center': 1, 'operations': 1, 'helicopters':
1, 'revelation': 1, 'outrage': 1, 'danger': 1, 'caterer': 1, 'reluctance': 1, 'manage
ment': 1, 'negligence': 1, 'foot': 1, 'tens': 1, 'thousands': 1, 'thanks': 1, 'billio
ns': 1, 'dollars': 1, 'funding': 1, 'access': 1, 'links': 1, 'aura': 1, 'competence':
1, 'week': 1, 'appearance': 1, 'crowd': 1, 'evidence': 1, 'affirmation': 1, 'camera
s': 1, 'babies': 1, 'selfies': 1, 'scene': 1, 'interactions': 1, 'pandemic': 1, 'clos
eness': 1, 'spectacle': 1, 'sign': 1, 'adulation': 1, 'measure': 1, 'mishandling': 1,
'bewilderment': 1, 'systems': 1, 'frailty': 1, 'inability': 1, 'safety': 1, 'capita
l': 1, 'failure': 1, 'enemies': 1, 'cursory': 1, 'look': 1, 'discussions': 1, 'specul
ation': 1, 'presence': 1, 'sympathizers': 1, 'back': 1, 'ROTTEN': 1, 'contract': 1,
'return': 1, 'freedom': 1, 'annexation': 1, 'value': 1, 'protection': 1, 'upsurge':
1, 'hunger': 1, 'right': 1, 'turmoil': 1, 'crackdowns': 1, 'liberals': 1, 'Polls': 1,
'jump': 1, 'cabinet': 1, 'governors': 1, 'parliament': 1, 'party': 1, 'passivity': 1,
'face': 1, 'problem': 1, 'signs': 1, 'ability': 1, 'responsibilities': 1, 'upscale':
1, 'suburb': 1, 'users': 1, 'deserts': 1, 'symbol': 1, 'era': 1, 'call': 1, 'salvag
e': 1, 'rise': 1, 'popularity': 1, 'capture': 1, 'Wagner': 1, 'month': 1, 'study': 1,
'group': 1, 'triumph': 1, 'climb': 1, 'beginning': 1, '158th': 1, 'brazen': 1, 'confr
ontation': 1, 'complaints': 1, 'ammunition': 1, 'crusader': 1, 'corruption': 1, 'pos
t': 1, 'man': 1, 'stripe': 1, 'color': 1, 'welcome': 1, 'residents': 1, 'disgruntleme
nt': 1, 'extent': 1, 'ease': 1, 'incredulity': 1, 'buildings': 1, 'blink': 1, 'eye':
1, 'frontline': 1, 'building': 1, 'troubles': 1, 'bleaker': 1, 'situations': 1, 'reje
ction': 1, 'result': 1, 'dissatisfaction': 1, 'prosecution': 1, 'Reactions': 1, 'incu
rsions': 1, 'paramilitaries': 1, 'spring': 1, 'fate': 1, 'desire': 1, 'talks': 1, 'wi
thdrawal': 1, 'respondents': 1, 'times': 1, 'fight': 1, 'pursuit': 1, 'survival': 1,
'faction': 1, 'enemy': 1, 'supporters': 1, 'op': 1, 'ed': 1, 'director': 1, 'result
s': 1, 'bolder': 1, 'shift': 1, 'attitudes': 1, 'prospect': 1, 'destabilization': 1,
'disorder': 1, 'reaction': 1, 'second': 1, 'managers': 1, 'corporations': 1, 'shell':
1, 'vulnerabilities': 1, 'quash': 1, 'complacency': 1, 'conviction': 1, 'decay': 1,
'dynamics': 1, 'incoherence': 1, 'jolt': 1, 'LINE': 1, 'camp': 1, 'service': 1, 'cons
ervatives': 1, 'correspondents': 1, 'TV': 1, 'screws': 1, 'traitors': 1, 'condition
s': 1, 'alternative': 1, 'suggestion': 1, 'mind': 1, 'portion': 1, 'crackdown': 1, 'm
ood': 1, 'distance': 1, 'lifestyles': 1, 'business': 1, 'room': 1, 'raft': 1, 'bill
s': 1, 'antiregime': 1, 'repression': 1, 'stiffening': 1, 'feeling': 1, 'politics':
1, 'conversation': 1, 'cohort': 1, 'likes': 1, 'secretary': 1, 'chair': 1, 'brand':
1, 'emphasis': 1, 'values': 1, 'importance': 1, 'family': 1, 'children': 1, 'bonds':
1, 'primacy': 1, 'rights': 1, 'men': 1, 'climate': 1, 'uniqueness': 1, 'mainstream':
1, 'aid': 1, 'exit': 1, 'strategy': 1, 'Leaders': 1, 'rhetoric': 1, 'gritty': 1, 'sta
r': 1, 'wanes': 1, 'yesterday': 1, 'technocrats': 1, 'adviser': 1, 'charge': 1, 'fal
l': 1, 'reconstruction': 1, 'territories': 1, 'lot': 1, 'areas': 1, 'courage': 1, 'di
ligence': 1, 'practitioners': 1, 'information': 1, 'contrast': 1, 'conspiracy': 1, 't
heories': 1, 'eruption': 1, 'volcano': 1, 'debris': 1, 'beneficiaries': 1, 'insurgenc
y': 1, 'occasion': 1, 'scores': 1, 'desk': 1, 'jockey': 1, 'leaders': 1, 'facts': 1,
'rant': 1, 'interview': 1, 'indispensability': 1, 'clash': 1, 'struggles': 1, 'busine
ssmen': 1, 'Chekists': 1, 'defeat': 1, 'attempt': 1, 'moves': 1, 'factor': 1, 'trut
h': 1, 'senses': 1, 'step': 1, 'understanding': 1, 'tendency': 1, 'alarm': 1, 'bell
s': 1, 'activists': 1, 'reforms': 1, 'dictatorship': 1, 'strategies': 1, 'necessity':
1, 'one': 1, 'crimes': 1, 'carnage': 1, 'system': 1, 'development': 1, 'solidifying':
1, 'realization': 1, 'inadequacies': 1, 'style': 1, 'ways': 1, 'miscalculations': 1,
'decision': 1, 'loss': 1, 'charges': 1, 'justice': 1, 'killings': 1, 'budget': 1, 'ex
penditures': 1, 'tool': 1, 'emotions': 1, 'beliefs': 1, 'advantage': 1, 'populace':
1, 'documents': 1, 'tales': 1, 'liberation': 1, 'irritation': 1, 'skills': 1, 'mood
s': 1, 'anger': 1, 'optimism': 1, 'views': 1, 'foes': 1, 'vision': 1, 'coherence': 1,
'thinking': 1, 'Factions': 1, 'analysts': 1, 'attempts': 1, 'militias': 1, 'panoply':
1, 'opposite': 1, 'aide': 1, 'bodyguard': 1, 'units': 1, 'subculture': 1, 'Cossacks':
1, 'move': 1, 'proliferation': 1, 'weaponry': 1, 'ministry': 1, 'purges': 1, 'success
or': 1, 'concentration': 1, 'prerogatives': 1, 'mission': 1, 'quagmire': 1, 'deputie
s': 1, 'senators': 1, 'approach': 1, 'impatience': 1, 'risk': 1, 'form': 1, 'accounta
bility': 1, 'responses': 1, 'antiestablishment': 1, 'agenda': 1, 'reverse': 1, 'fract
```

```
iousness': 1, 'influence': 1, 'role': 1, 'changes': 1, 'circumstances': 1, 'directio
n': 1, 'act': 1, 'destiny': 1, 'fulfillment': 1, 'script': 1})

PROPN Counter({'Putin': 53, 'Prigozhin': 37, 'Russia': 27, 'Kremlin': 24, 'Russians':
23, 'Ukraine': 20, 'Moscow': 12, 'Wagner': 12, 'Shoigu': 8, 'May': 7, 'June': 7, 'Ukr
ainians': 6, 'West': 5, 'Rostov': 5, 'Levada': 5, 'Patrushev': 5, 'Don': 4, 'Belgoro
d': 3, 'Alexander': 3, 'Defense': 3, 'Ministry': 3, 'Security': 3, 'Medvedev': 3, 'Zo
lotov': 3, 'Rosgvardia': 3, 'President': 2, 'Crimea': 2, 'Bakhmut': 2, 'Sergey': 2,
'Rublyovka': 2, 'Foreign': 2, 'Minister': 2, 'Ordinary': 2, 'FSB': 2, 'Sergei': 2, 'C
ouncil': 2, 'Dmitry': 2, 'Kyiv': 2, 'Americans': 2, 'Vladimir': 1, 'February': 1, 'Kh
arkiv': 1, 'September': 1, 'October': 1, 'Kerch': 1, 'Strait': 1, 'Yevgeny': 1, 'Luka
shenko': 1, 'FRAGILE': 1, 'STATE': 1, 'Victory': 1, 'Day': 1, 'Red': 1, 'Square': 1,
'Ermochenko': 1, 'Reuters': 1, 'Valery': 1, 'Gerasimov': 1, 'Dagestan': 1, 'COVID-1
9': 1, 'United': 1, 'Yeltsin': 1, 'Center': 1, 'Romir': 1, 'Lavrov': 1, 'Prime': 1,
'Mikhail': 1, 'Mishustin': 1, 'Facebook': 1, 'Federal': 1, 'Services': 1, 'Forbes':
1, 'Denis': 1, 'Volkov': 1, 'HARDENING': 1, 'Bastrykin': 1, 'Investigative': 1, 'Comm
ittee': 1, 'Narysh\xadkin': 1, 'Intelligence': 1, 'Service': 1, 'Nikolai': 1, 'Vyache
slav': 1, 'Volodin': 1, 'State': 1, 'Duma': 1, 'Putinism': 1, 'Kiriyenko': 1, 'Mara
t': 1, 'Khusnullin': 1, 'Anglo': 1, 'Saxon': 1, 'U.S.': 1, 'Yellowstone': 1, 'Nationa
l': 1, 'Park': 1, 'August': 1, 'Evgenia': 1, 'Novozhenina': 1, 'Viktor': 1, 'app': 1,
'Telegram': 1, 'SCRIPT': 1, 'Mironov': 1, 'Andrei': 1, 'Troshev': 1})
```

The printout shows the most common words, grouped by part of speech; for verbs the most common is 'become', for nouns the most common is 'war', etc.

1. Find a subject/object relationship through the dependency parser in any sentence.

```
In [12]:   # Dependency Parser
           def pr_tree(word, level):
               if word.is_punct:
                   return
               for child in word.lefts:
                   pr_tree(child,level+1)
               print('\t' * level + word.text + ' - ' + word.dep_)
               for child in word.rights:
                   pr_tree(child, level+1)
```

```
In [13]:   # Run Parser over article text:
           iteration_count = 0
           for sentence in article_text.sents:
               pr_tree(sentence.root, 0)
               print('--------------------------')

               iteration_count += 1
               if iteration_count >= 2:
                   break
```

```
                    After - mark
                                Russian - amod
                        President - compound
                        Vladimir - compound
                    Putin - nsubj
            launched - ccomp
                            an - det
                    invasion - dobj
                            of - prep
                                    Ukraine - pobj
                    in - prep
                            February - pobj
                                    2022 - nummod
                    Russian - amod
            elites - nsubj
acted - ROOT
                    as - mark
                    if - mark
                            the - det
                    war - nsubj
                    had - aux
                    not - neg
                    really - advmod
            changed - advcl
                    anything - dobj
                    on - prep
                                    the - det
                                    home - compound
                        front - pobj
--------------------------
                    Even - advmod
                    as - mark
                            the - det
                    campaign - nsubj
            foundered - advcl
                    and - cc
                            the - det
                                    West - npadvmod
                        tightened - amod
                    sanctions - conj
                        on - prep
                                        the - det
                                        Russian - amod
                                economy - pobj
            those - nsubj
                    with - prep
                            power - pobj
                                    in - prep
                                            Moscow - pobj
seemed - ROOT
                    to - aux
            carry - xcomp
                    on - prt
                    as - prep
                            usual - amod
--------------------------
```

The first subject-object relationship in the article is that between 'Putin' (subject). and 'invasion' (object). Note: The above loop can run over the whole article, but makes the resulting output

extremely long; I have included a 'break' after 2 iterations.

1. Show the most common Entities and their types.

In [14]:
```python
iteration_count = 0
for entity in article_text.ents:
    print(entity, entity.label_)

    iteration_count += 1

    if iteration_count >= 20:
        break
```

```
Russian NORP
Vladimir Putin PERSON
Ukraine GPE
February 2022 DATE
Russian NORP
Russian NORP
Moscow GPE
last autumn DATE
Ukrainian NORP
September 2022 DATE
Russian NORP
Kremlin ORG
October DATE
Ukrainian GPE
the Kerch Strait LOC
Crimea ORG
Russia GPE
Kremlin ORG
months DATE
Russia GPE
```

The most common entities are 'Russian' (Entity type: Nationality or religious/political group), Vladimir Putin (Entity type: person), and Ukraine (Entity type: Geo-Political). Note: I again included a break to keep the printout manageable. To avoid repeating this note, I did the same for several below parts of the workbook.

1. Find Entities and their dependency (hint: entity.root.head)

In [15]:
```python
iteration_count = 0
for entity in article_text.ents:
    print(entity,";", entity.root.head)

    iteration_count += 1

    if iteration_count >= 20:
        break
```

```
Russian ; President
Vladimir Putin ; launched
Ukraine ; of
February 2022 ; in
Russian ; elites
Russian ; economy
Moscow ; in
last autumn ; Since
Ukrainian ; counterattack
September 2022 ; in
Russian ; positions
Kremlin ; launched
October ; in
Ukrainian ; strike
the Kerch Strait ; bridge
Crimea ; between
Russia ; Crimea
Kremlin ; redlines
months ; prior
Russia ; war
```

The above lists the entities and their dependencies.

1. Find the most similar words in the article

```
In [16]:  # Initialize Distinct Token Set
          distinct_tokens = set()

          # Creating a Set of All Tokens (i.e. Words) From the Article
          for token in article_text:
              distinct_tokens.add(token.text)
```

```
In [17]:  # Cleaning the set to remove uninteresting tokens (page breaks, punctuation, stop word
          cleaned_tokens = [token.lower() for token in distinct_tokens if token not in punctuati
```

```
In [18]:  # Token Similarity Comparison Function (with functiionality to save the most similar s
          highest_similarity_scores = []
          base_word = []

          def similarity_finder_save_highest(word, article_text):
              # Processor applied to word
              word_token = processor(word)

              # Copy of Article for this iteration of loop
              article_tokens = article_text.copy()

              # Initialize Similarity Score List (for this iteration)
              similarity_scores = []

              for token in article_tokens:
                  # Skipping Matching Words to avoid 100% similarities
                  if token.text.lower() == word.lower():
                      continue

                  similarity = word_token.similarity(token)
                  similarity_scores.append((token.text, similarity))
              # Add to list of highest similarities
```

```python
        highest_similarity_scores.append(max(similarity_scores, key=lambda x: x[1]))
        base_word.append(word)
        # Return Scores for Later Access
        return similarity_scores, highest_similarity_scores
```

In [19]:
```python
# Loop Through Cleaned Tokens with the similarity finder function
for token in cleaned_tokens:
    similarity_finder_save_highest(token, article_text)
```

C:\Users\Brett\AppData\Local\Temp\ipykernel_25944\2180956133.py:20: UserWarning: [W00
7] The model you're using has no word vectors loaded, so the result of the Doc.simila
rity method will be based on the tagger, parser and NER, which may not give useful si
milarity judgements. This may happen if you're using one of the small models, e.g. `e
n_core_web_sm`, which don't ship with word vectors and only use context-sensitive ten
sors. You can always add your own word vectors, or use one of the larger models inste
ad if available.
  similarity = word_token.similarity(token)

In [20]:
```python
# Convert to dataframe for ease of viewing & sorting
base_word_dataframe = pd.DataFrame(base_word)
similar_word_dataframe = pd.DataFrame(highest_similarity_scores)
similarities_dataframe = pd.concat([base_word_dataframe, similar_word_dataframe], axis
new_column_names = ['Base_Word', 'Comparitive_word','Similarity_Score']
similarities_dataframe.columns = new_column_names
similarities_dataframe.sort_values(by='Similarity_Score', ascending=False, inplace=Tru
```

In [21]:
```python
similarities_dataframe
```

Out[21]:

	Base_Word	Comparitive_word	Similarity_Score
1314	calling	complicating	0.822883
54	maintaining	seizing	0.821239
1153	discussing	complicating	0.821223
956	directing	complicating	0.811001
693	deepening	seizing	0.809884
...	...	...	...
886	volodin	Leaders	0.407490
925	facto	obsolete	0.405063
1059	dozen	.	0.388313
110	hardcore	quickly	0.380046
946	mikhail	.	0.359952

1490 rows × 3 columns

The most similar tokens are 'calling' and 'complicating', showing a 0.822 similarity score.