

DETECTING LINGUISTIC & RHETORICAL CHANGE ON REDDIT

Vector Space Embeddings, Content Analysis 2020

Adarsh Mathew

OVERVIEW & DATA SOURCES

- Right-wing subreddits and the 2016 election
- An alt-right takeover of the *r/The_Donald* subreddit: Move from traditional Conservative rhetoric to alt-right rhetoric?
 - Immigration
 - Media 'bias'
 - Traditional
 - Social Media
- 3 years of data from select subreddits:
 - *r/The_Donald*, *r/Incels*, *r/altright*
 - Start: May 2015
 - Announced candidacy in Republican primary
 - End: Nov 2018

OVERVIEW & DATA SOURCES

- Right-wing subreddits and the 2016 election
- An alt-right takeover of the *r/The_Donald* subreddit: Move from traditional Conservative rhetoric to alt-right rhetoric?
 - Immigration
 - Media 'bias'
 - Traditional
 - Social Media
- 3 years of data from select subreddits:
 - *r/The_Donald*, *r/Incels*, *r/altright*
 - Start: May 2015
 - Announced candidacy in Republican primary
 - End: Nov 2018

```
donald4_W2V.most_similar('news')
```

```
[('fox', 0.9715726375579834),  
 ('report', 0.943863034248352),  
 ('cnn', 0.9353094100952148),  
 ('carly', 0.9350003004074097),  
 ('scandal', 0.9348690509796143),  
 ('national', 0.9188006520271301),  
 ('i', 0.916765570640564),  
 ('rule', 0.9153780937194824),  
 ('pac', 0.9090359210968018),  
 ('fund', 0.9084845781326294)]
```

```
donald8_W2V.most_similar('news')
```

```
[('fake', 0.958344578742981),  
 ('source', 0.9509509205818176),  
 ('video', 0.9501924514770508),  
 ('twitter', 0.9485650658607483),  
 ('article', 0.9461296796798706),  
 ('specialist', 0.9417068958282471),  
 ('compilation', 0.9410412311553955),  
 ('tweet', 0.9401676058769226),  
 ('obscure', 0.9395379424095154),  
 ('pizza', 0.9386619925498962)]
```

NEWS – FAKE OR FOX?

- T4, Super Tuesday (GOP Nomination):
 - Popular media outlets closely associated with ‘news’
- T8, General Election month:
 - ‘Fake news’
 - Twitter makes an appearance, Fox and CNN don’t show up
 - ‘Pizzagate’ reference

```
donald4_W2V.most_similar('news')
```

```
[('fox', 0.9715726375579834),  
 ('report', 0.943863034248352),  
 ('cnn', 0.9353094100952148),  
 ('carly', 0.9350003004074097),  
 ('scandal', 0.9348690509796143),  
 ('national', 0.9188006520271301),  
 ('i', 0.916765570640564),  
 ('rule', 0.9153780937194824),  
 ('pac', 0.9090359210968018),  
 ('fund', 0.9084845781326294)]
```

```
donald8_W2V.most_similar('news')
```

```
[('fake', 0.958344578742981),  
 ('source', 0.9509509205818176),  
 ('video', 0.9501924514770508),  
 ('twitter', 0.9485650658607483),  
 ('article', 0.9461296796798706),  
 ('specialist', 0.9417068958282471),  
 ('compilation', 0.9410412311553955),  
 ('tweet', 0.9401676058769226),  
 ('obscure', 0.9395379424095154),  
 ('pizza', 0.9386619925498962)]
```

NEWS – FAKE OR FOX?

- T4, Super Tuesday (GOP Nomination):
 - Popular media outlets closely associated with ‘news’
- T8, General Election month:
 - ‘Fake news’
 - Twitter makes an appearance, Fox and CNN don’t show up
 - ‘Pizzagate’ reference

```
donald4_W2V.most_similar('mexican')
```

```
[('u.s', 0.9978862404823303),  
 ('refugee', 0.9968644380569458),  
 ('citizen', 0.9962786436080933),  
 ('china', 0.995557427406311),  
 ('islamic', 0.9953213334083557),  
 ('system', 0.994718074798584),  
 ('trade', 0.9943752884864807),  
 ('workforce', 0.9938507080078125),  
 ('exploit', 0.9937943816184998),  
 ('undocumented', 0.9934136867523193)]
```

```
donald8_W2V.most_similar('mexican')
```

```
[('ideology', 0.9991565346717834),  
 ('young', 0.9990911483764648),  
 ('tear', 0.9990442395210266),  
 ('sjw', 0.9988831281661987),  
 ('ass', 0.9988289475440979),  
 ('disgrace', 0.9988203048706055),  
 ('jewish', 0.998818039894104),  
 ('wall', 0.9988049864768982),  
 ('smart', 0.9987630844116211),  
 ('table', 0.9987449645996094)]
```

MEXICAN – IMMIGRATION VS IDENTITY POLITICS

- T4, Super Tuesday (GOP Nomination):
 - Discourse seems economic
- T8, General Election month:
 - Discourse seems more closely related to identity & ideology
 - Ideology, SJW, Jewish (?)

```
donald4_W2V.most_similar('mexican')
```

```
[('u.s', 0.9978862404823303),  
 ('refugee', 0.9968644380569458),  
 ('citizen', 0.9962786436080933),  
 ('china', 0.995557427406311),  
 ('islamic', 0.9953213334083557),  
 ('system', 0.994718074798584),  
 ('trade', 0.9943752884864807),  
 ('workforce', 0.9938507080078125),  
 ('exploit', 0.9937943816184998),  
 ('undocumented', 0.9934136867523193)]
```

```
donald8_W2V.most_similar('mexican')
```

```
[('ideology', 0.9991565346717834),  
 ('young', 0.9990911483764648),  
 ('tear', 0.9990442395210266),  
 ('sjw', 0.9988831281661987),  
 ('ass', 0.9988289475440979),  
 ('disgrace', 0.9988203048706055),  
 ('jewish', 0.998818039894104),  
 ('wall', 0.9988049864768982),  
 ('smart', 0.9987630844116211),  
 ('table', 0.9987449645996094)]
```

MEXICAN – IMMIGRATION VS IDENTITY POLITICS

- T4, Super Tuesday (GOP Nomination):
 - Discourse seems economic
- T8, General Election month:
 - Discourse seems more closely related to identity & ideology
 - Ideology, SJW, Jewish (?)

```
donald4_W2V.most_similar('immigration')
```

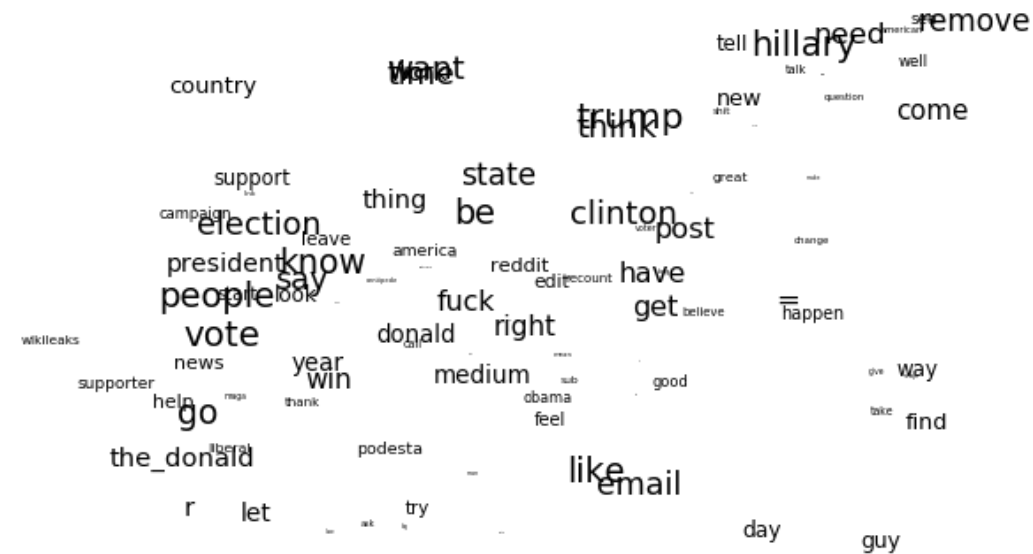
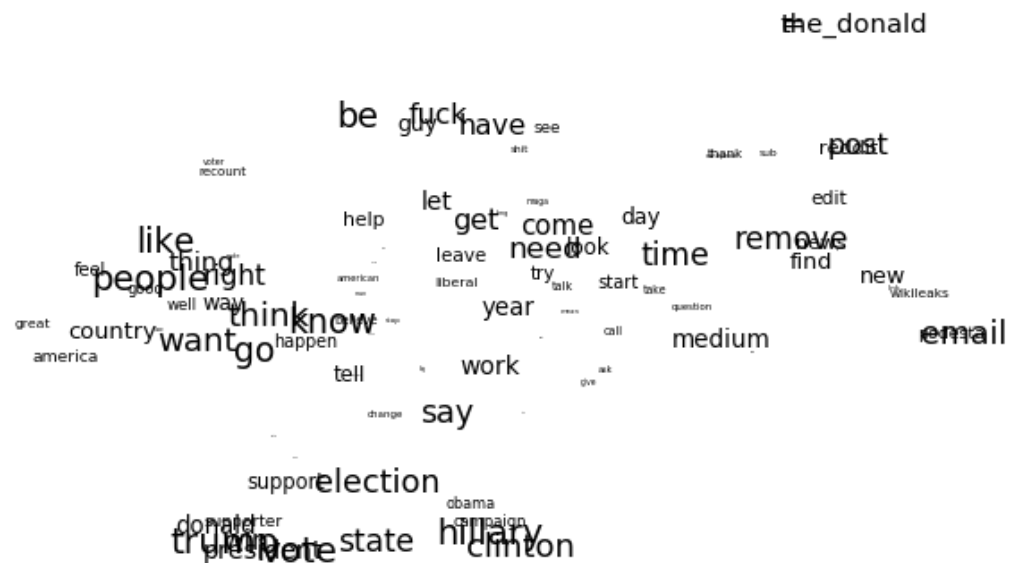
```
[('immigrant', 0.9980612397193909),  
 ('illegal', 0.984207034111023),  
 ('border', 0.9784513711929321),  
 ('american', 0.9715807437896729),  
 ('alien', 0.9578444361686707),  
 ('mexico', 0.9573886394500732),  
 ('government', 0.9465535879135132),  
 ('law', 0.9439253807067871),  
 ('trade', 0.9404299259185791),  
 ('wall', 0.9388926029205322)]
```

```
donald8_W2V.most_similar('immigration')
```

```
[('bear', 0.9997624158859253),  
 ('division', 0.9996041059494019),  
 ('protection', 0.9995772838592529),  
 ('pull', 0.9995088577270508),  
 ('intend', 0.9994912147521973),  
 ('terrorism', 0.9994733333587646),  
 ('son', 0.9994708299636841),  
 ('healthcare', 0.9994666576385498),  
 ('mistake', 0.999466061592102),  
 ('ship', 0.9994590282440186)]
```

IMMIGRATION – ECONOMIC VS SECURITY

- T4, Super Tuesday (GOP Nomination):
 - Discourse seems around illegal immigration, trade, and the wall
- T8, General Election month:
 - More ‘fear’-related words: division, protection, terrorism, healthcare, mistake



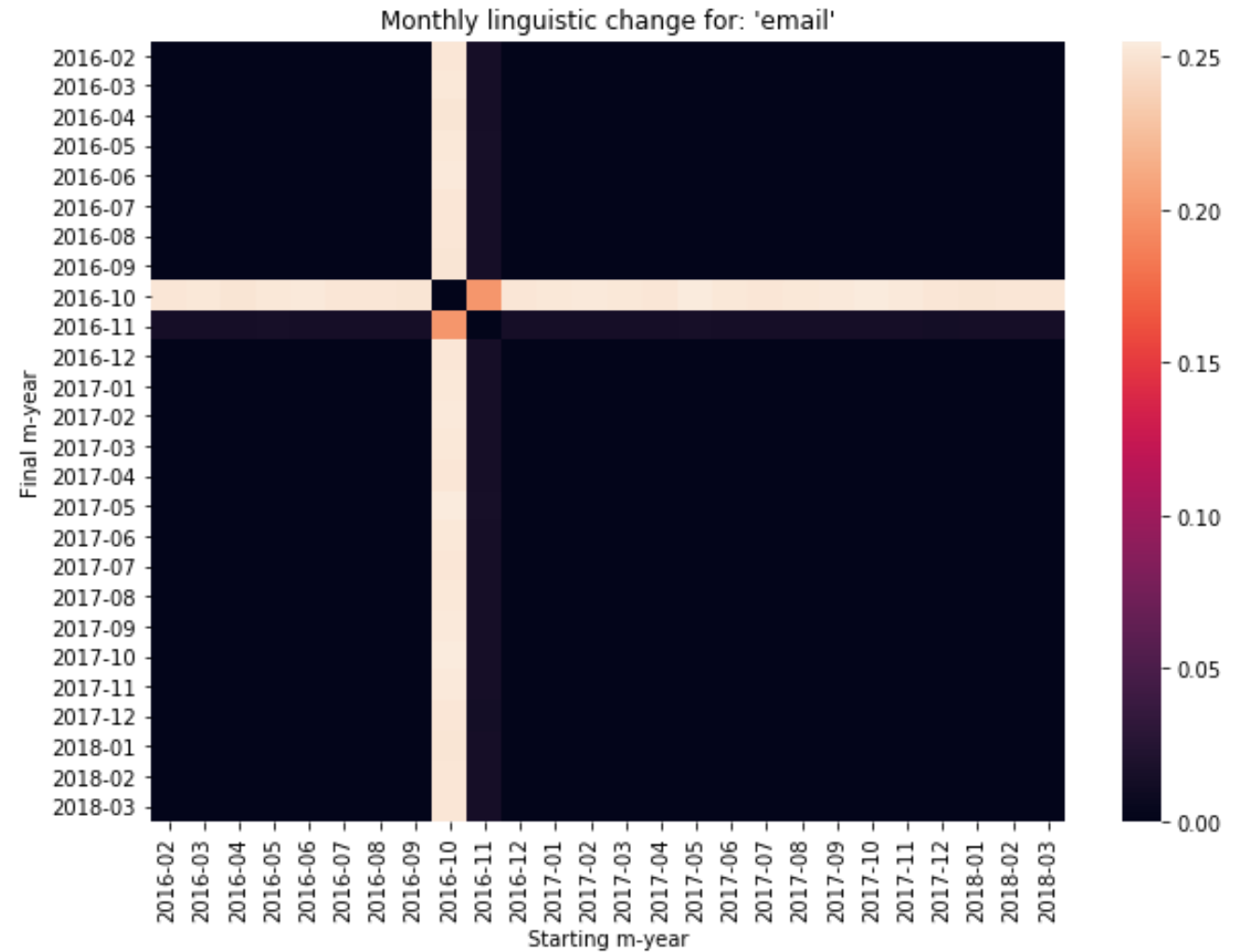
TSNE WORD DISTRIBUTIONS

EX4B: LINGUISTIC CHANGE/DIFFERENCE

- Adapted from Hamilton, Jurafsky, Lescovec paper
- Three key functions:
 - ``intersection_align_gensim``:
 - Intersects two word2vec models.
 - Keeps shared vocab only. Does it lose out on new words that enter the corpus?
 - ``Procrustes_align_gensim``: Aligns Model 2 to Model 1 after intersection
 - ``getDivergence``:
 - Pairwise Cosine Dissimilarity of a word for a category (time, true class, topic, etc.)

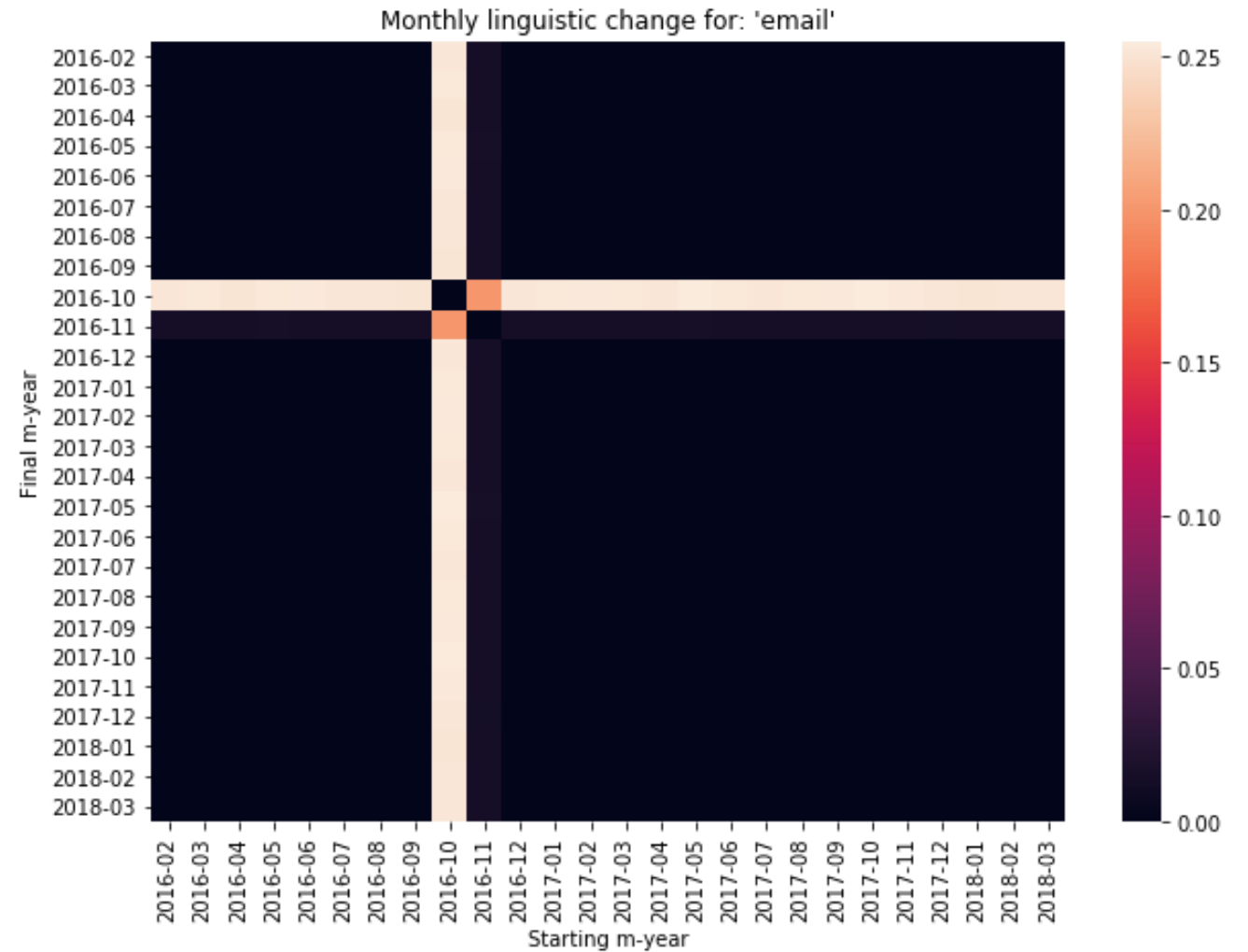
CLINTON'S
EMAILS
MATTERED*

- Vastly different meaning in October 2016 and reversion to old meaning right after
- Wedge issue only in that one month
- Oversized reportage



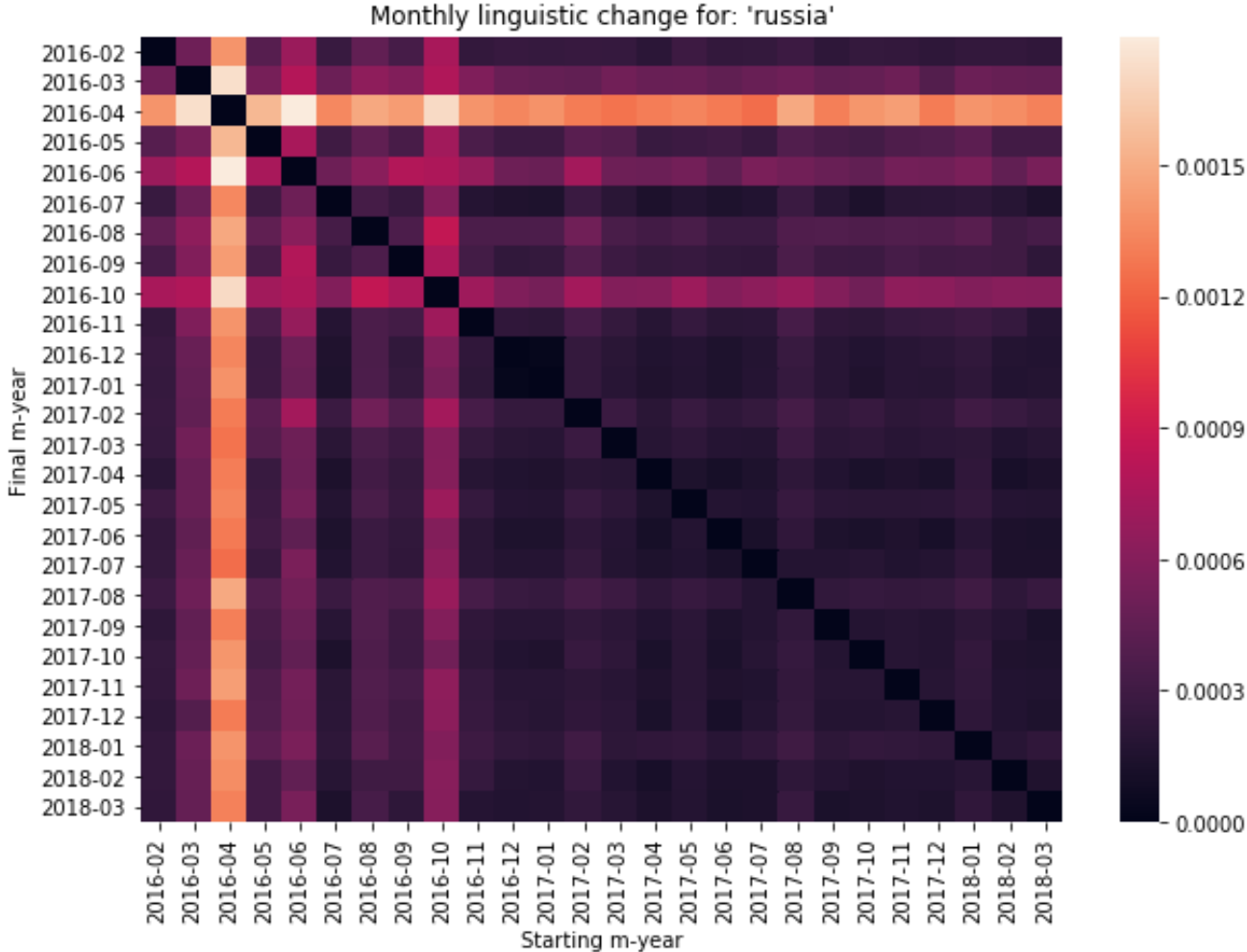
CLINTON'S
EMAILS
MATTERED*

- Vastly different meaning in October 2016 and reversion to old meaning right after
- Wedge issue only in that one month
- Oversized reportage



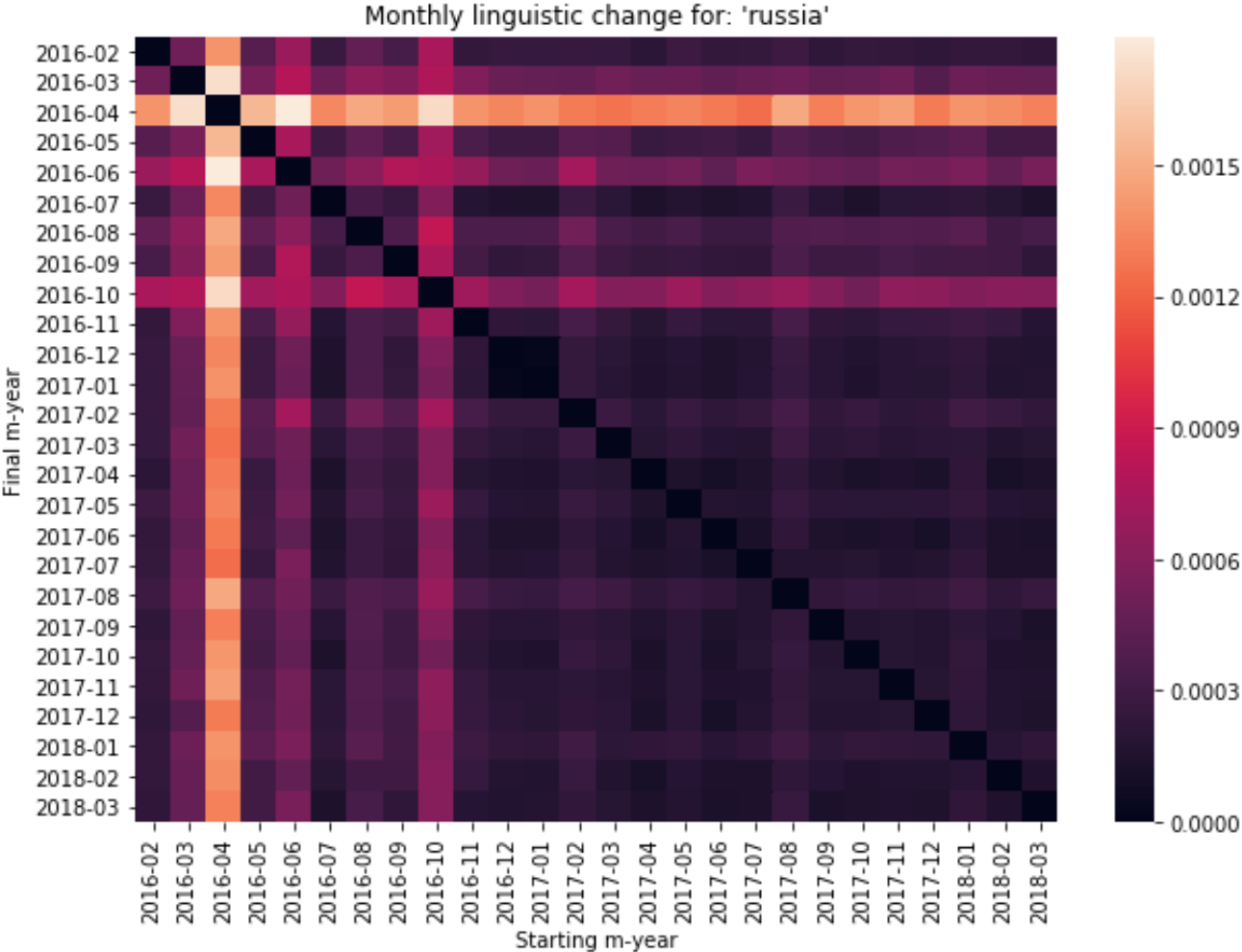
CHANGING NATURE OF 'RUSSIA'

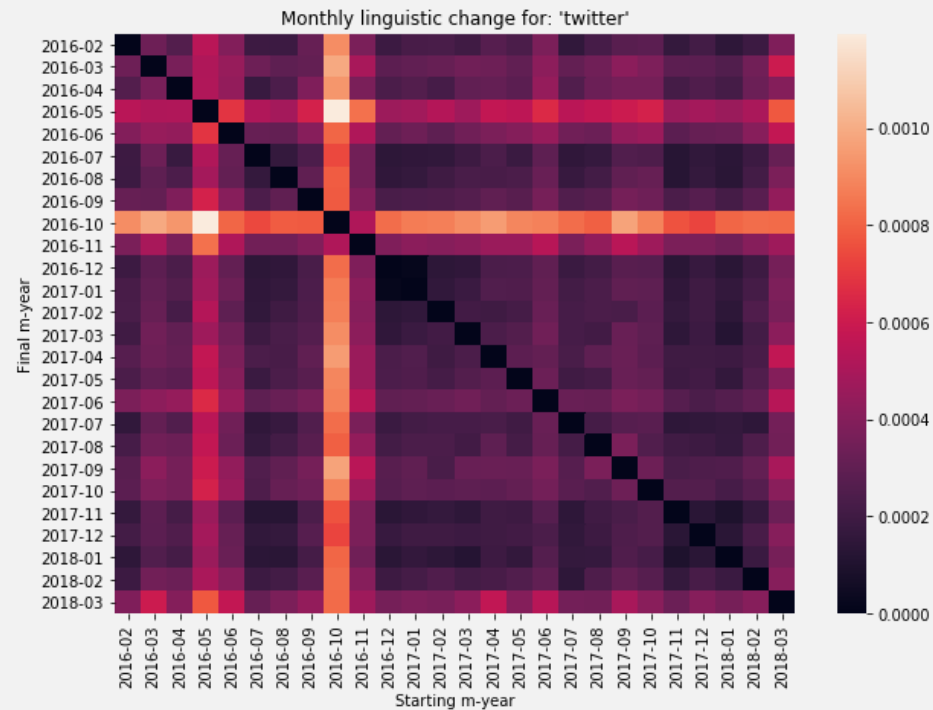
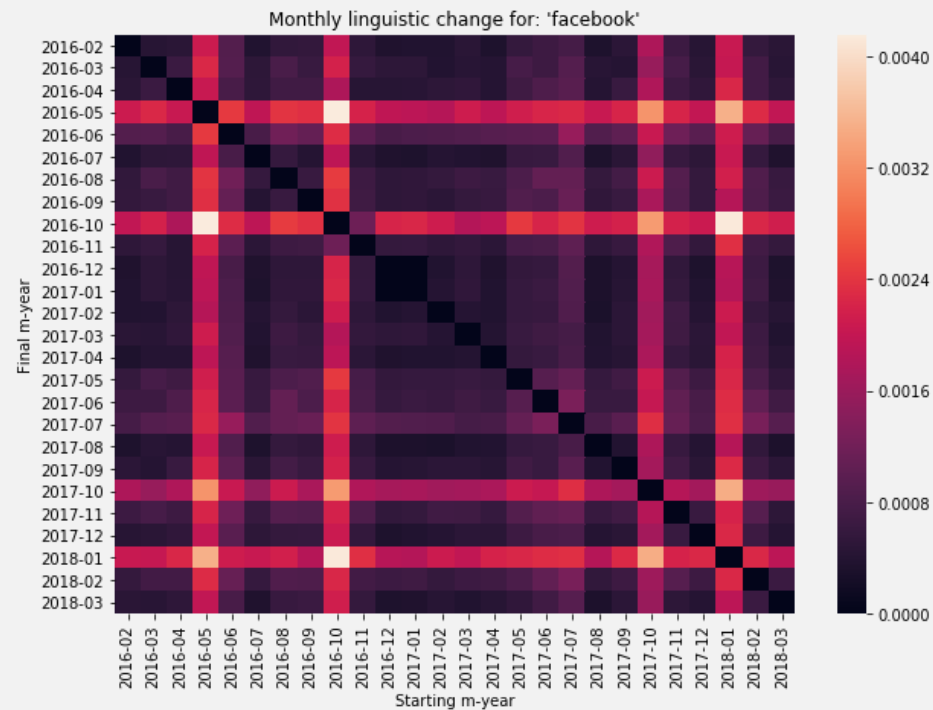
- April 2016: First news of DNC hack
- October 2016: Election month – email hack takes centre-stage
- After October 2016: Minimal change => some consensus about Russia's role? (even if the interpretation may vary)



CHANGING NATURE OF 'RUSSIA'

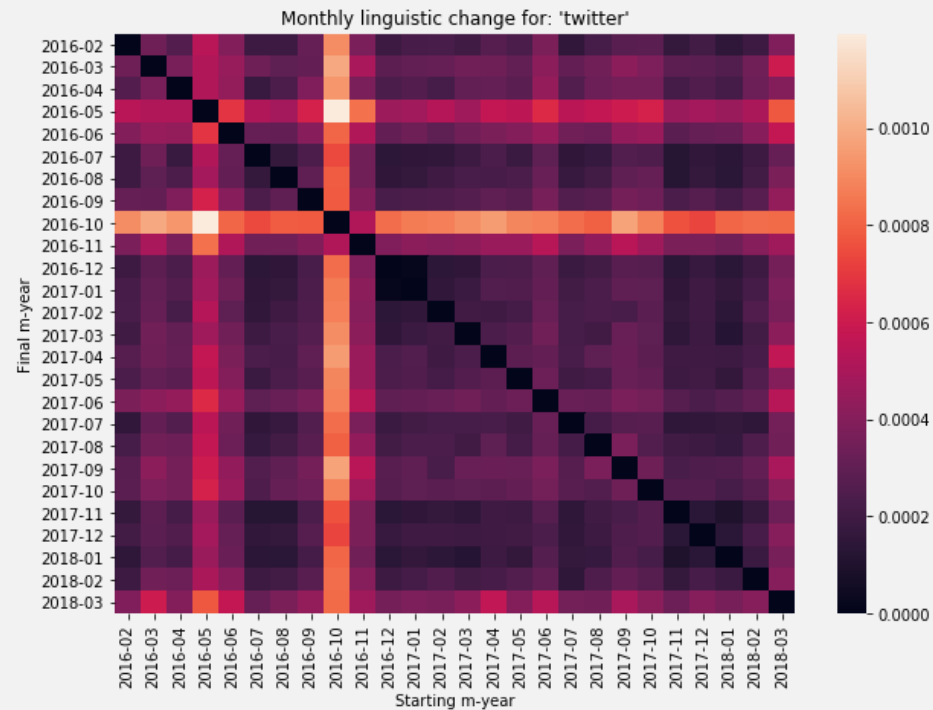
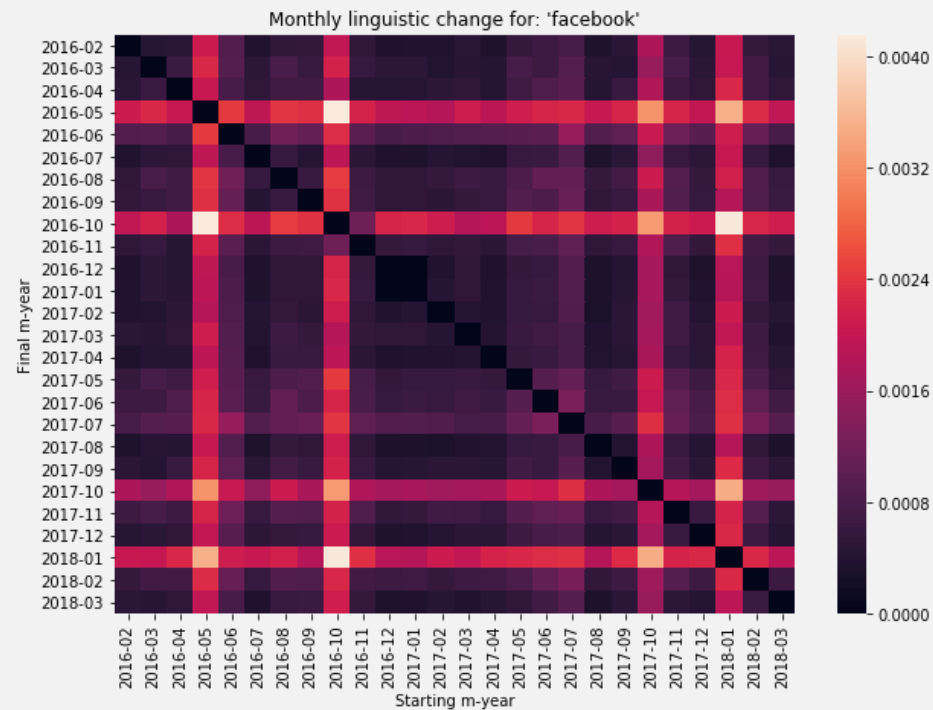
- April 2016: First news of DNC hack
- October 2016: Election month – email hack takes centre-stage
- After October 2016: Minimal change => some consensus about Russia's role? (even if the interpretation may vary)





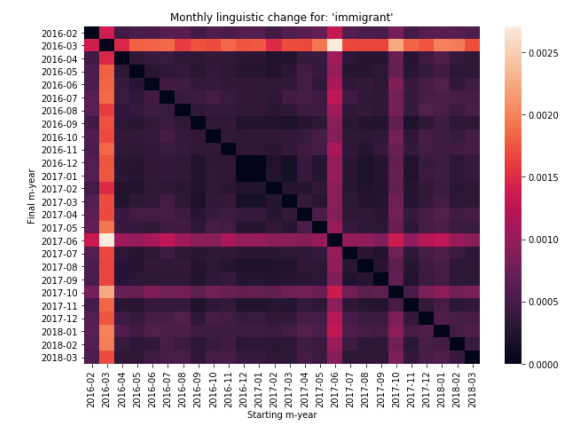
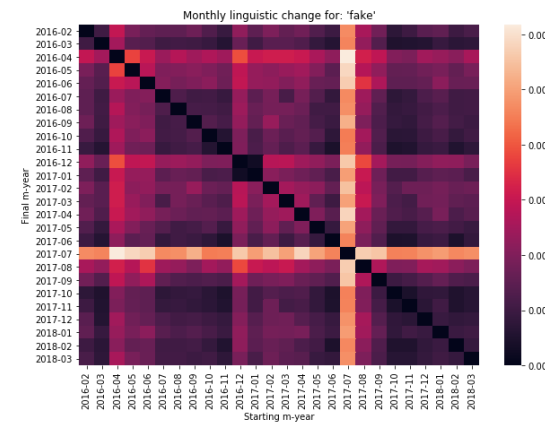
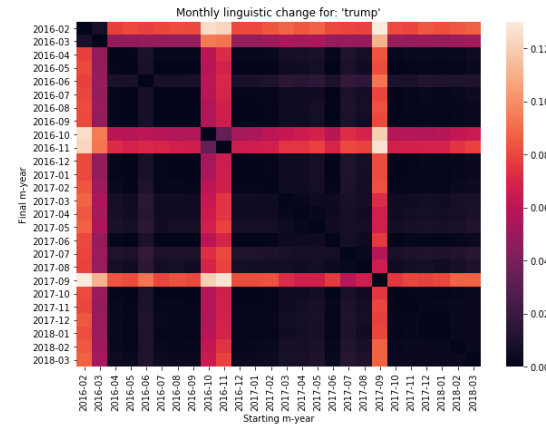
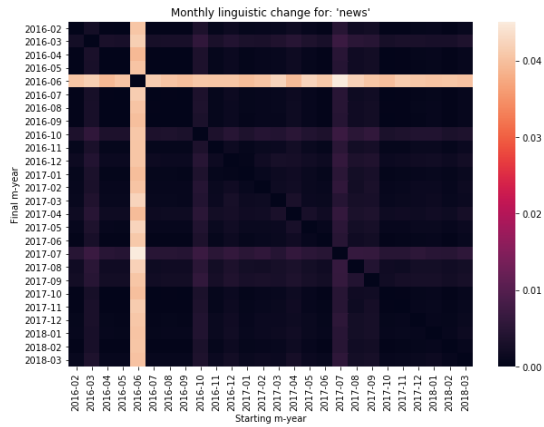
ROLE OF FACEBOOK

- Greater change in the linguistic context surrounding Facebook over Twitter
 - Higher scale
 - Starker changes
 - Trump's Twitter usage as unsurprising
- Early 2018: Cambridge Analytica scandal

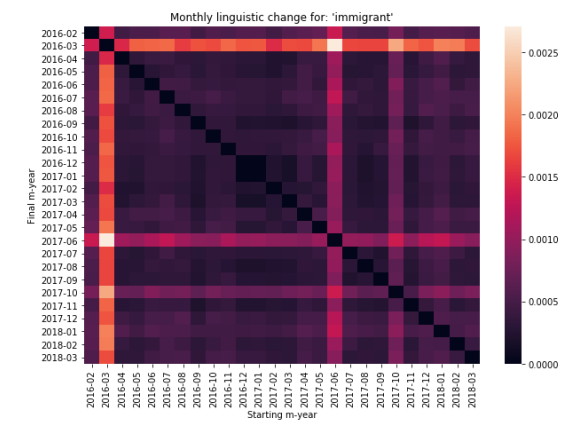
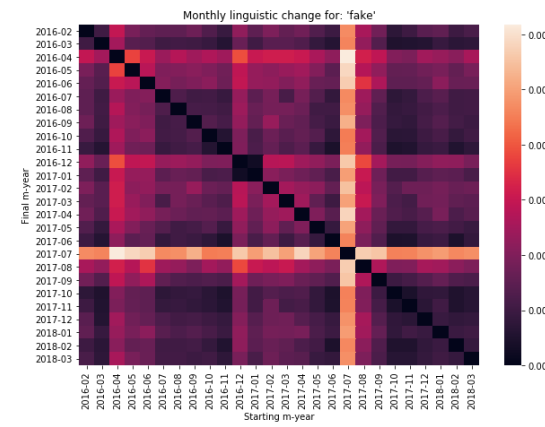
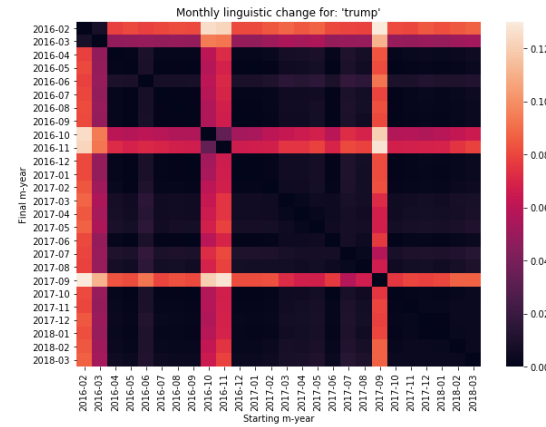
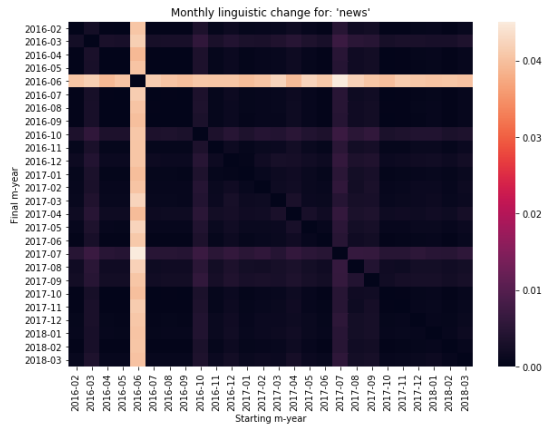


ROLE OF FACEBOOK

- Greater change in the linguistic context surrounding Facebook over Twitter
 - Higher scale
 - Starker changes
 - Trump's Twitter usage as unsurprising
- Early 2018: Cambridge Analytica scandal



SELECTIVE STABILITY?



SELECTIVE STABILITY?

NEXT STEPS

- Good initial results, but how to make the narrative of linguistic/rhetorical change more robust?
 - Larger time scales for each embedding? (3 month-period)
 - Event-driven analysis? (Super Tuesday, October 2016, Charlottesville, Muslim ban)
- How to account for change in vocabulary over time?
- Complete subreddit discussions (submission + comments) as documents for Doc2Vec?

NEXT STEPS

- Good initial results, but how to make the narrative of linguistic/rhetorical change more robust?
 - Larger time scales for each embedding? (3 month-period)
 - Event-driven analysis? (Super Tuesday, October 2016, Charlottesville, Muslim ban)
- How to account for change in vocabulary over time?
- Complete subreddit discussions (submission + comments) as documents for Doc2Vec?