Exploring the use of syntactic superlatives on Twitter

Data

Counts for *'de/het meest *e'* (syntactic superlative) and *'de/het *ste'* (morphological superlative) per month for 2013-2020 Dutch twitter data.

For the morphological data, we filtered occurrences of *'eerste, laatste, meeste, juiste, beste'* as these are very frequent (approx. 60% of the cases) while the syntactic alternative does not really exist.

Note that we did not filter retweets, this might have some effect especially for counts of specific word forms.

Loading the raw counts:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
setwd("C:/Users/gosse/haytabo/CorpusLinguistics")
meest <- read.csv("meest_super.csv", sep=' ')
meestt <- tibble(meest)
meestt
```

```
## # A tibble: 96 x 3
##    Month      meest  super
##    <chr>      <int>  <int>
##  1 2013-01-01 29152 309411
##  2 2013-02-01 11746 129540
##  3 2013-03-01 28450 290426
##  4 2013-04-01 25002 251980
##  5 2013-05-01 27577 296507
##  6 2013-06-01 23970 261582
##  7 2013-07-01 28919 311402
##  8 2013-08-01 27505 302790
##  9 2013-09-01  8770  98693
## 10 2013-10-01 27897 277235
## # ... with 86 more rows
```

Adding some rows

Adding rows for total and ratio of meest/super and percentage of *meest* in total

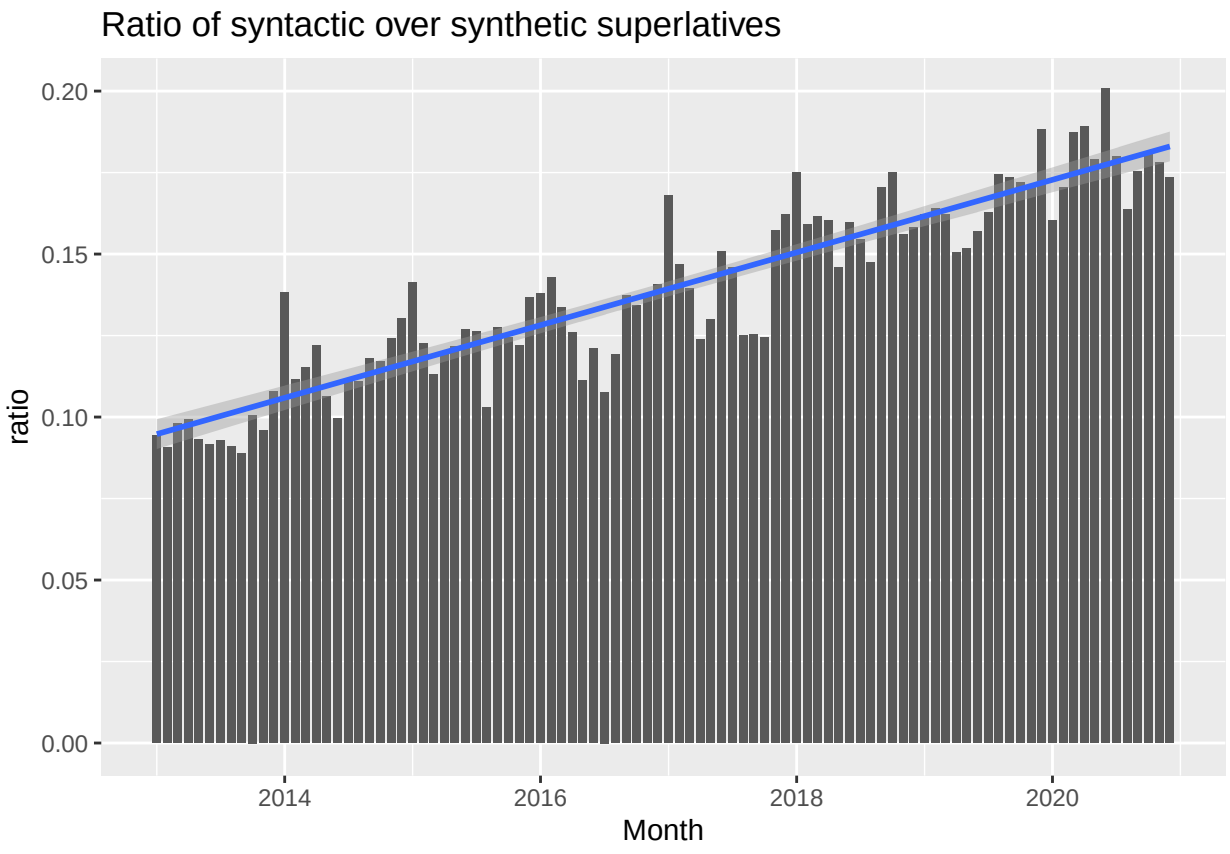
```
meestt2 <- mutate(meestt, total = meest + super, ratio = meest/super, percentage = meest/total)
meestt2
```

```
## # A tibble: 96 x 6
##    Month      meest  super  total  ratio percentage
##    <chr>      <int>  <int>  <int>  <dbl>      <dbl>
##  1 2013-01-01 29152 309411 338563 0.0942     0.0861
##  2 2013-02-01 11746 129540 141286 0.0907     0.0831
##  3 2013-03-01 28450 290426 318876 0.0980     0.0892
##  4 2013-04-01 25002 251980 276982 0.0992     0.0903
##  5 2013-05-01 27577 296507 324084 0.0930     0.0851
##  6 2013-06-01 23970 261582 285552 0.0916     0.0839
##  7 2013-07-01 28919 311402 340321 0.0929     0.0850
##  8 2013-08-01 27505 302790 330295 0.0908     0.0833
##  9 2013-09-01  8770  98693 107463 0.0889     0.0816
## 10 2013-10-01 27897 277235 305132 0.101      0.0914
## # ... with 86 more rows
```

Plotting the trend

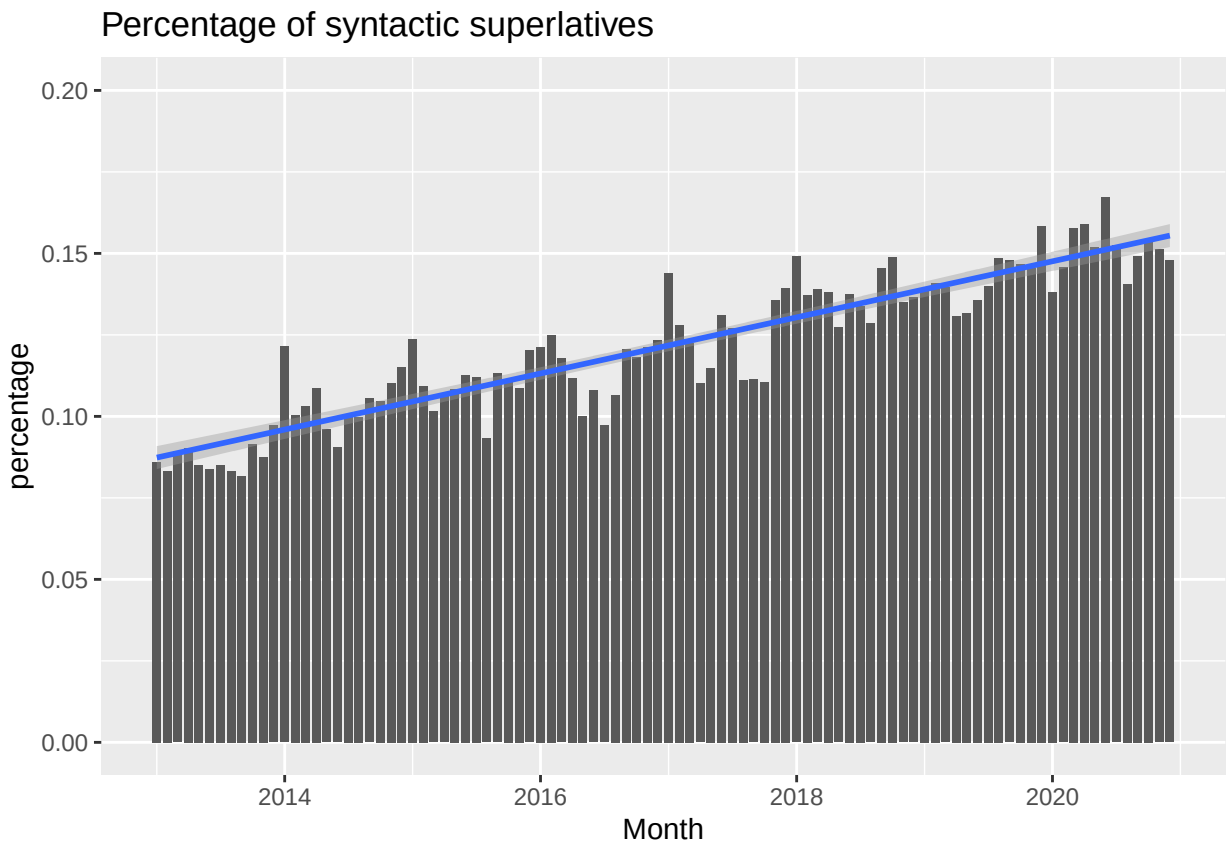
```
library(ggplot2)

ggplot(data = meestt2, aes(x=as.Date(Month), y = ratio)) + geom_col() + geom_smooth(method = "lm") +
    coord_cartesian(ylim = c(0.0,0.2)) + xlab('Month') + ggtitle("Ratio of syntactic over synthetic sup
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(data = meestt2, aes(x=as.Date(Month), y = percentage)) + geom_col() + geom_smooth(method = "lm")
    coord_cartesian(ylim = c(0.0,0.2)) + xlab('Month') + ggtitle("Percentage of syntactic superlatives"
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Individual forms

Does this trend also hold when we check for individual forms. Here we try *'meest populaire'* vs *'populairste'*.

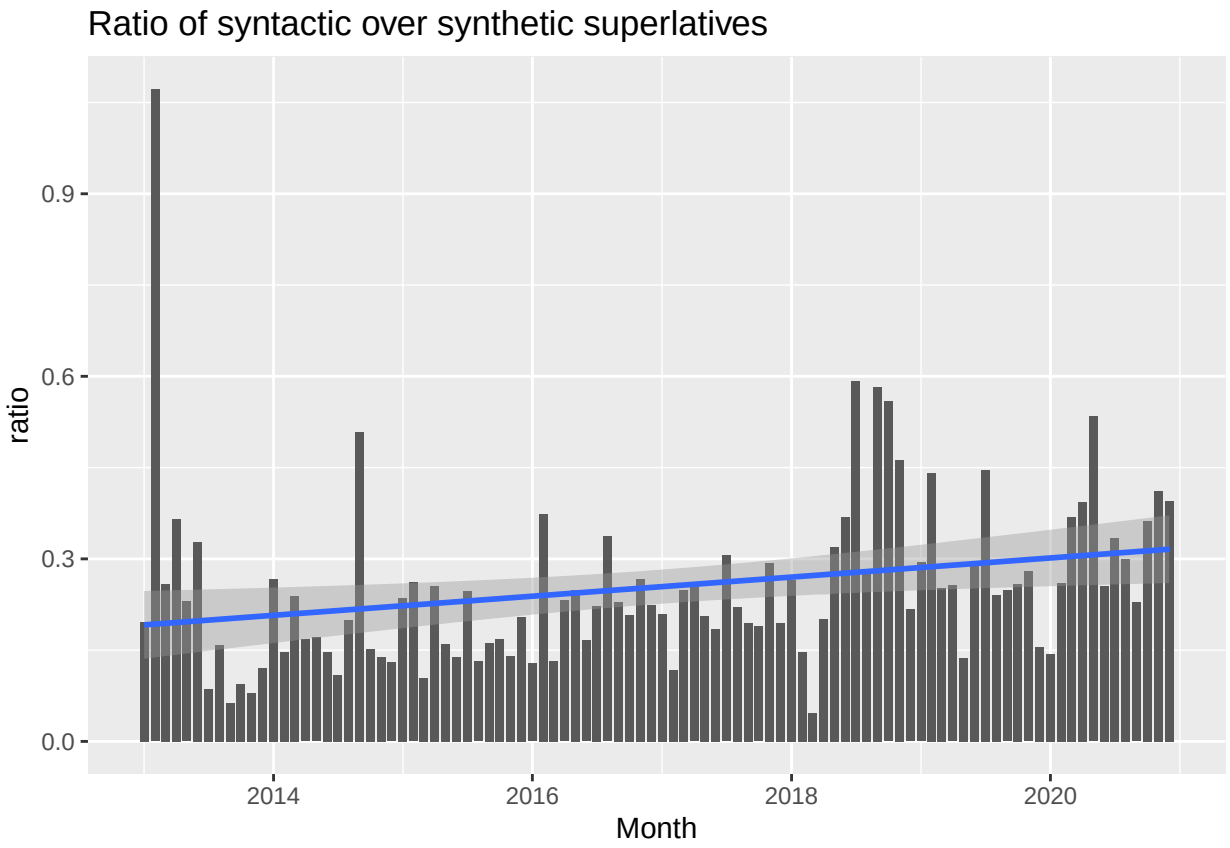
It seems the trend is in the same direction, but the data is quite a bit more noisy, due to lower total numbers.

```
populair <- read.csv("meest_populaire.csv", sep=' ')
populairr <- tibble(populair)
populairr2 <- mutate(populairr, total = meest + super, ratio = meest/super, percentage = meest/total)
populairr2
```

```
## # A tibble: 96 x 6
##    Month      meest super total  ratio percentage
##    <chr>      <int> <int> <int>  <dbl>      <dbl>
##  1 2013-01-01   311  1580  1891 0.197      0.164
##  2 2013-02-01   330   308   638 1.07       0.517
##  3 2013-03-01   231   891  1122 0.259      0.206
##  4 2013-04-01   257   702   959 0.366      0.268
##  5 2013-05-01   202   879  1081 0.230      0.187
##  6 2013-06-01   206   629   835 0.328      0.247
##  7 2013-07-01   218  2541  2759 0.0858     0.0790
##  8 2013-08-01   247  1554  1801 0.159      0.137
##  9 2013-09-01    35   552   587 0.0634     0.0596
## 10 2013-10-01   201  2125  2326 0.0946     0.0864
## # ... with 86 more rows
```

```
ggplot(data = populairr2, aes(x=as.Date(Month), y = ratio)) + geom_col() + geom_smooth(method = "lm") +
    xlab('Month') + ggtitle("Ratio of syntactic over synthetic superlatives")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Prefix aller-

The prefix *aller-* is quite popular in the superlative as well (*allerliefste*). Does it increase in popularity as well? Here we compare counts for *aller-* prefixed superlatives with all superlatives.

The trend is downward. (Quick hypothesis: as Twitter is being used more and more for professional communication, informal language decreases as well. We see this with other phenomena as well, such as frequency of spelling errors.)

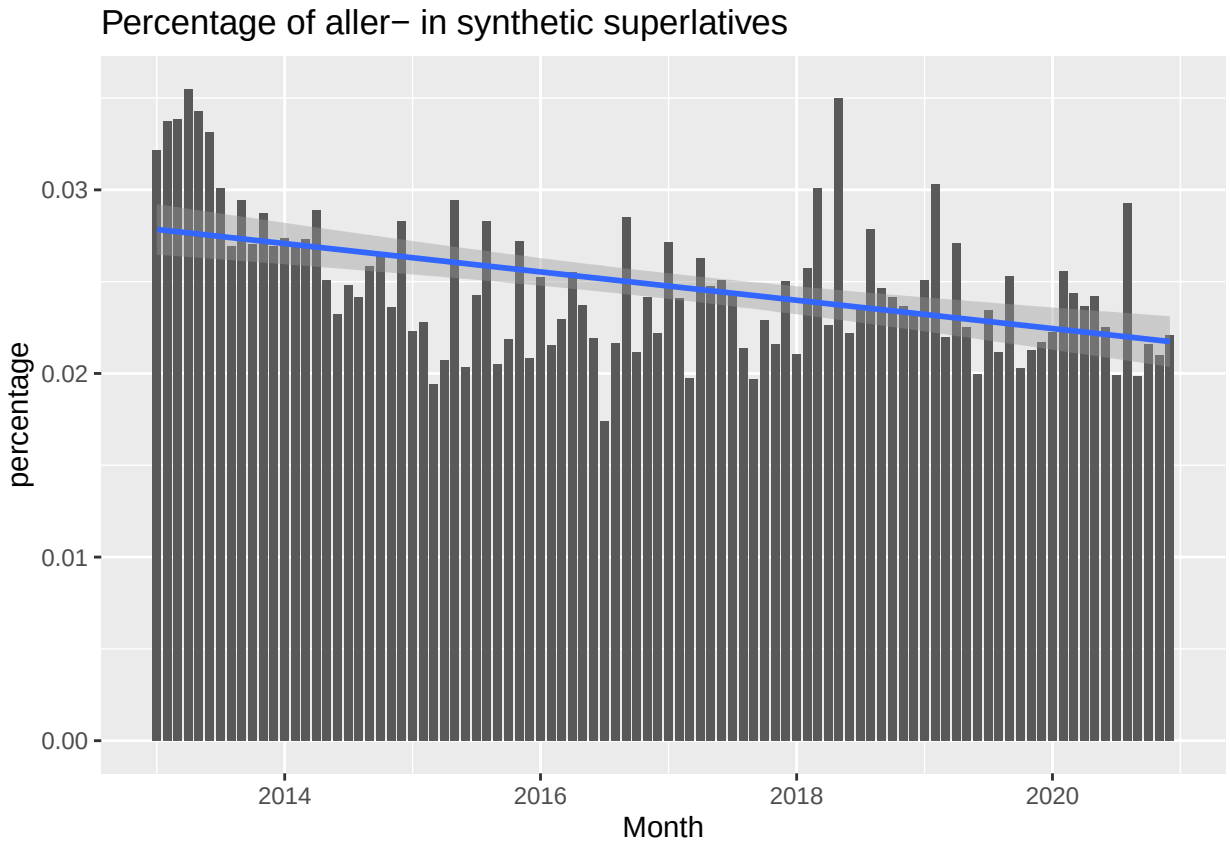
```
aller <- read.csv("aller.csv", sep=' ')
allerr <- tibble(aller)
allerr2 <- mutate(allerr,  percentage = aller/super)
allerr2
```

```
## # A tibble: 96 x 4
##    Month      aller  super percentage
##    <chr>      <int>  <int>      <dbl>
##  1 2013-01-01  9949 309411     0.0322
##  2 2013-02-01  4368 129540     0.0337
##  3 2013-03-01  9833 290426     0.0339
##  4 2013-04-01  8947 251980     0.0355
##  5 2013-05-01 10173 296507     0.0343
```

```
##  6 2013-06-01  8669 261582     0.0331
##  7 2013-07-01  9373 311402     0.0301
##  8 2013-08-01  8148 302790     0.0269
##  9 2013-09-01  2903  98693     0.0294
## 10 2013-10-01  7492 277235     0.0270
## # ... with 86 more rows
```

```
ggplot(data = allerr2, aes(x=as.Date(Month), y = percentage)) + geom_col() + geom_smooth(method = "lm") +
    xlab('Month') + ggtitle("Percentage of aller- in synthetic superlatives")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



‘‘‘