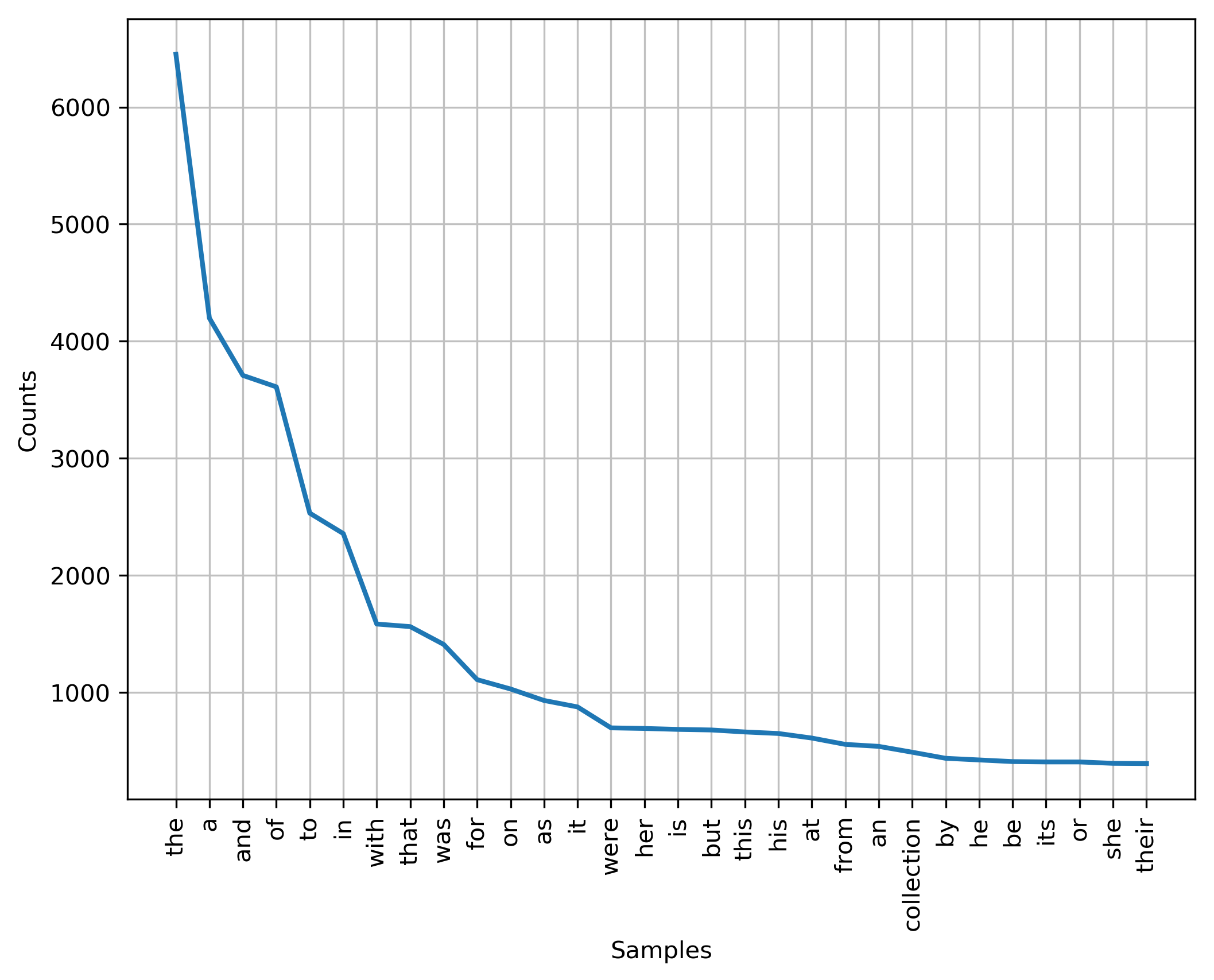
A report in word document. In the report, you need to present the results of 4 approaches (in terms of frequency and plot). Compare the results, and provide your insights about

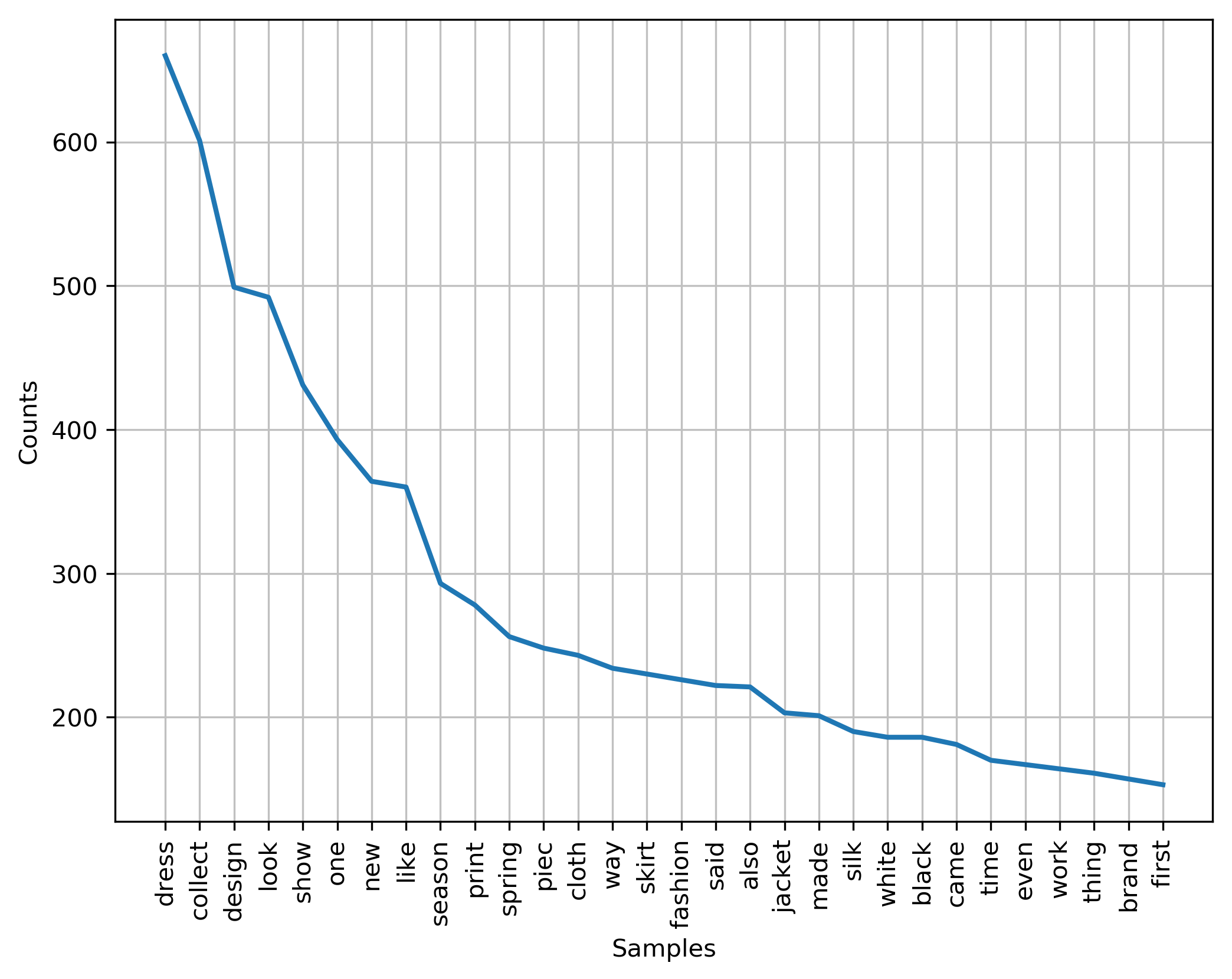
1. Simple approach：

|  |  |
| --- | --- |
| Word | Frequency |
| the | 6451 |
| a | 4197 |
| and | 3708 |
| of | 3610 |
| to | 2531 |
| in | 2357 |
| with | 1584 |
| that | 1562 |
| was | 1410 |
| for | 1109 |
| on | 1029 |
| as | 931 |
| it | 876 |
| were | 697 |
| her | 692 |
| is | 684 |
| but | 679 |
| this | 662 |
| his | 649 |
| at | 610 |
| from | 556 |
| an | 539 |
| collection | 489 |
| by | 437 |
| he | 423 |
| be | 409 |
| its | 406 |
| or | 406 |
| she | 394 |
| their | 392 |



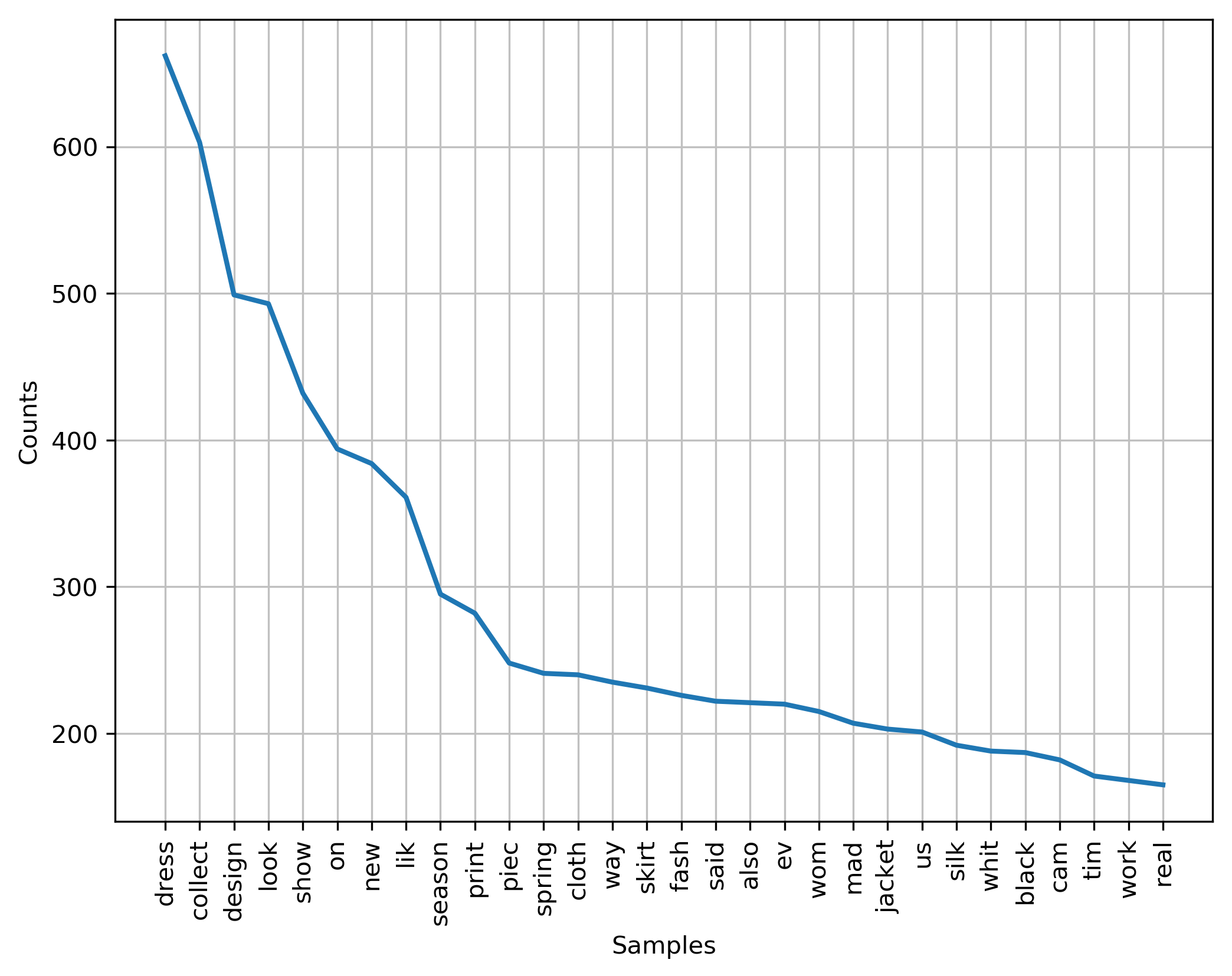
1. a bag-of-words approach with stemming and stop words removal
   1. Porter

|  |  |
| --- | --- |
| **Word** | **Frequency** |
| dress | 660 |
| collect | 601 |
| design | 499 |
| look | 492 |
| show | 431 |
| one | 393 |
| new | 364 |
| like | 360 |
| season | 293 |
| print | 278 |
| spring | 256 |
| piec | 248 |
| cloth | 243 |
| way | 234 |
| skirt | 230 |
| fashion | 226 |
| said | 222 |
| also | 221 |
| jacket | 203 |
| made | 201 |
| silk | 190 |
| white | 186 |
| black | 186 |
| came | 181 |
| time | 170 |
| even | 167 |
| work | 164 |
| thing | 161 |
| brand | 157 |
| first | 153 |



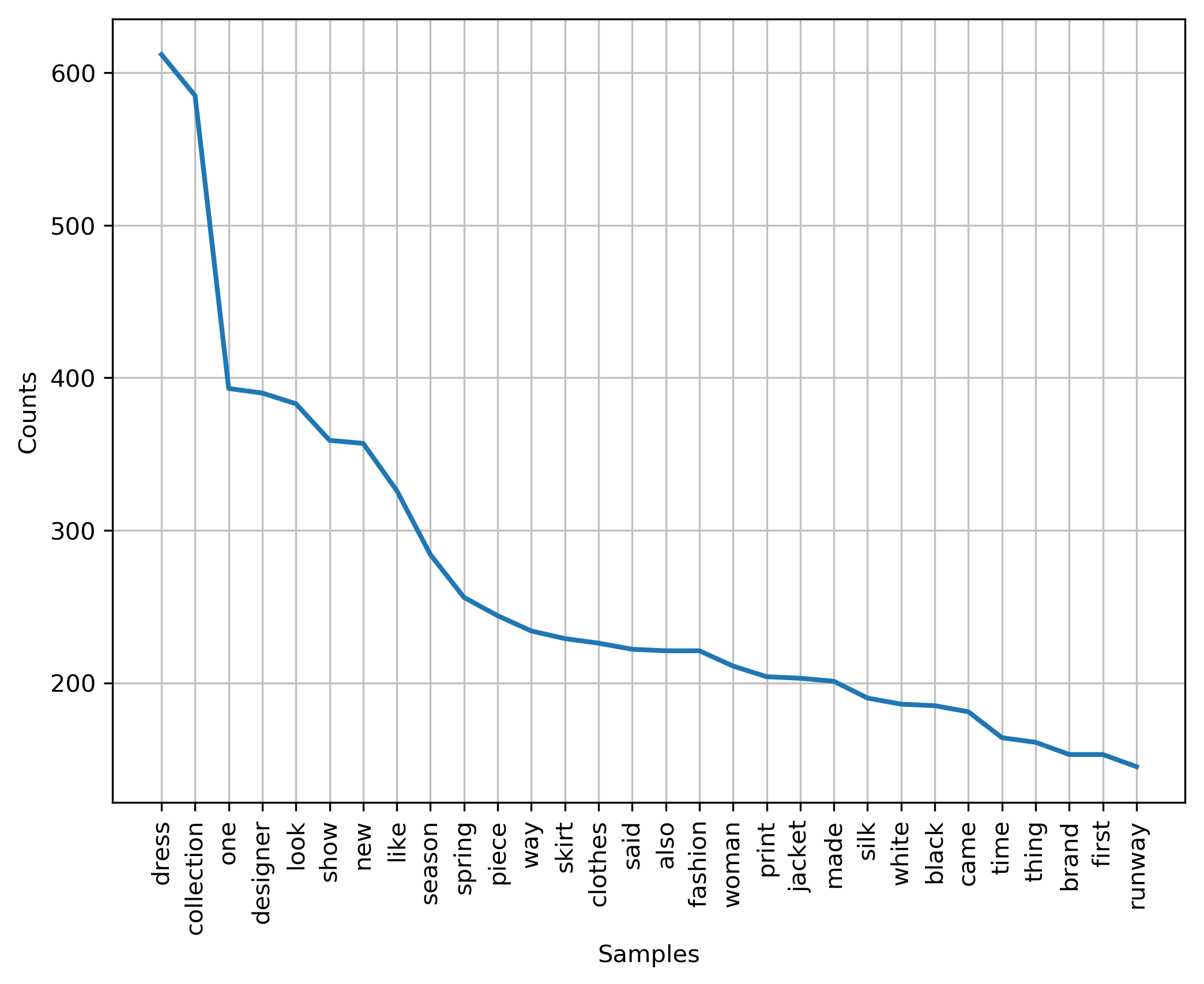
* 1. Lancaster

|  |  |
| --- | --- |
| Word | Frequency |
| dress | 662 |
| collect | 603 |
| design | 499 |
| look | 493 |
| show | 432 |
| on | 394 |
| new | 384 |
| lik | 361 |
| season | 295 |
| print | 282 |
| piec | 248 |
| spring | 241 |
| cloth | 240 |
| way | 235 |
| skirt | 231 |
| fash | 226 |
| said | 222 |
| also | 221 |
| ev | 220 |
| wom | 215 |
| mad | 207 |
| jacket | 203 |
| us | 201 |
| silk | 192 |
| whit | 188 |
| black | 187 |
| cam | 182 |
| tim | 171 |
| work | 168 |
| real | 165 |



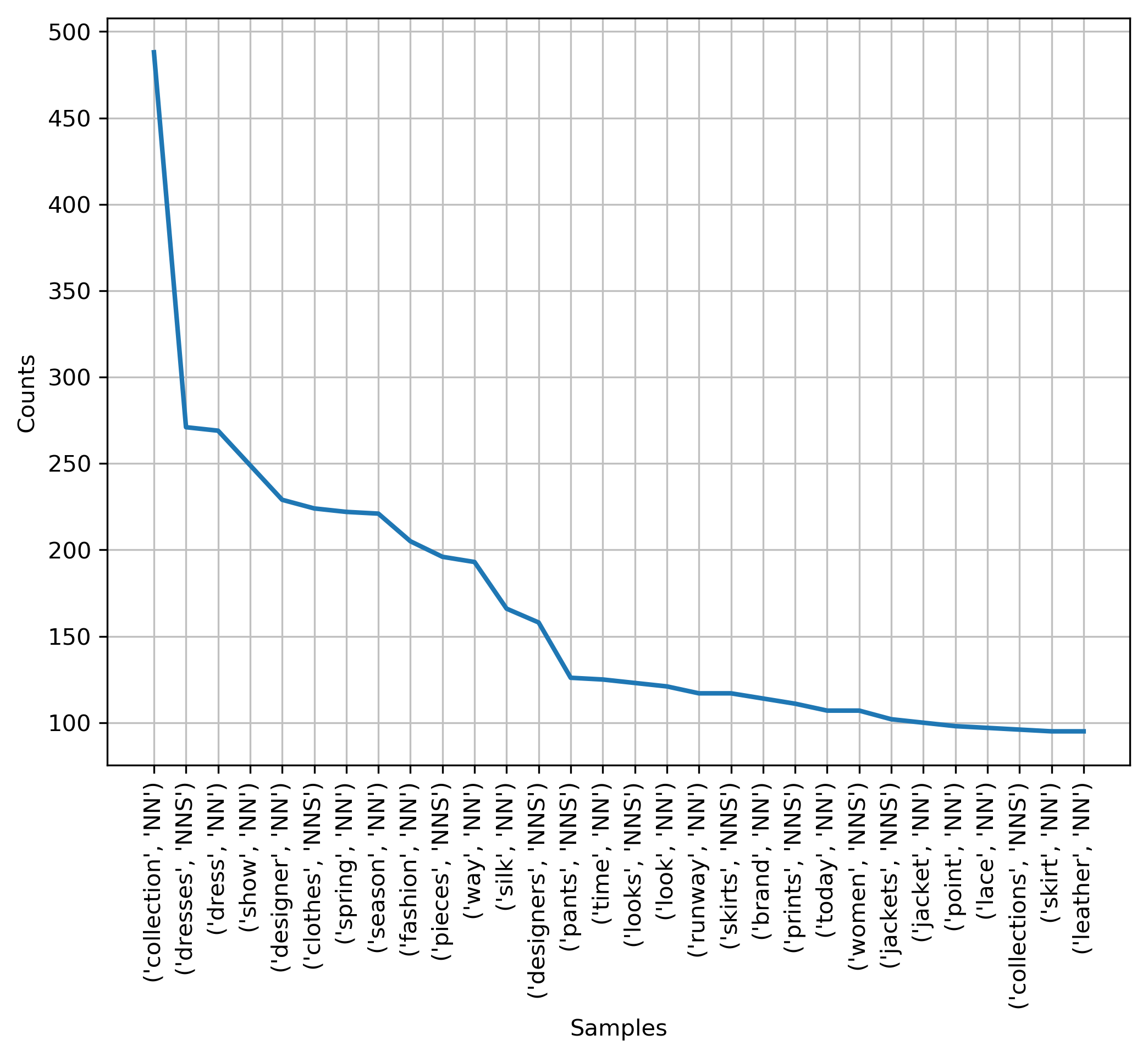
* 1. WorldNet Lemmatizer

|  |  |
| --- | --- |
| **Word** | **Frequency** |
| dress | 612 |
| collection | 585 |
| one | 393 |
| designer | 390 |
| look | 383 |
| show | 359 |
| new | 357 |
| like | 326 |
| season | 284 |
| spring | 256 |
| piece | 244 |
| way | 234 |
| skirt | 229 |
| clothes | 226 |
| said | 222 |
| also | 221 |
| fashion | 221 |
| woman | 211 |
| print | 204 |
| jacket | 203 |
| made | 201 |
| silk | 190 |
| white | 186 |
| black | 185 |
| came | 181 |
| time | 164 |
| thing | 161 |
| brand | 153 |
| first | 153 |
| runway | 145 |



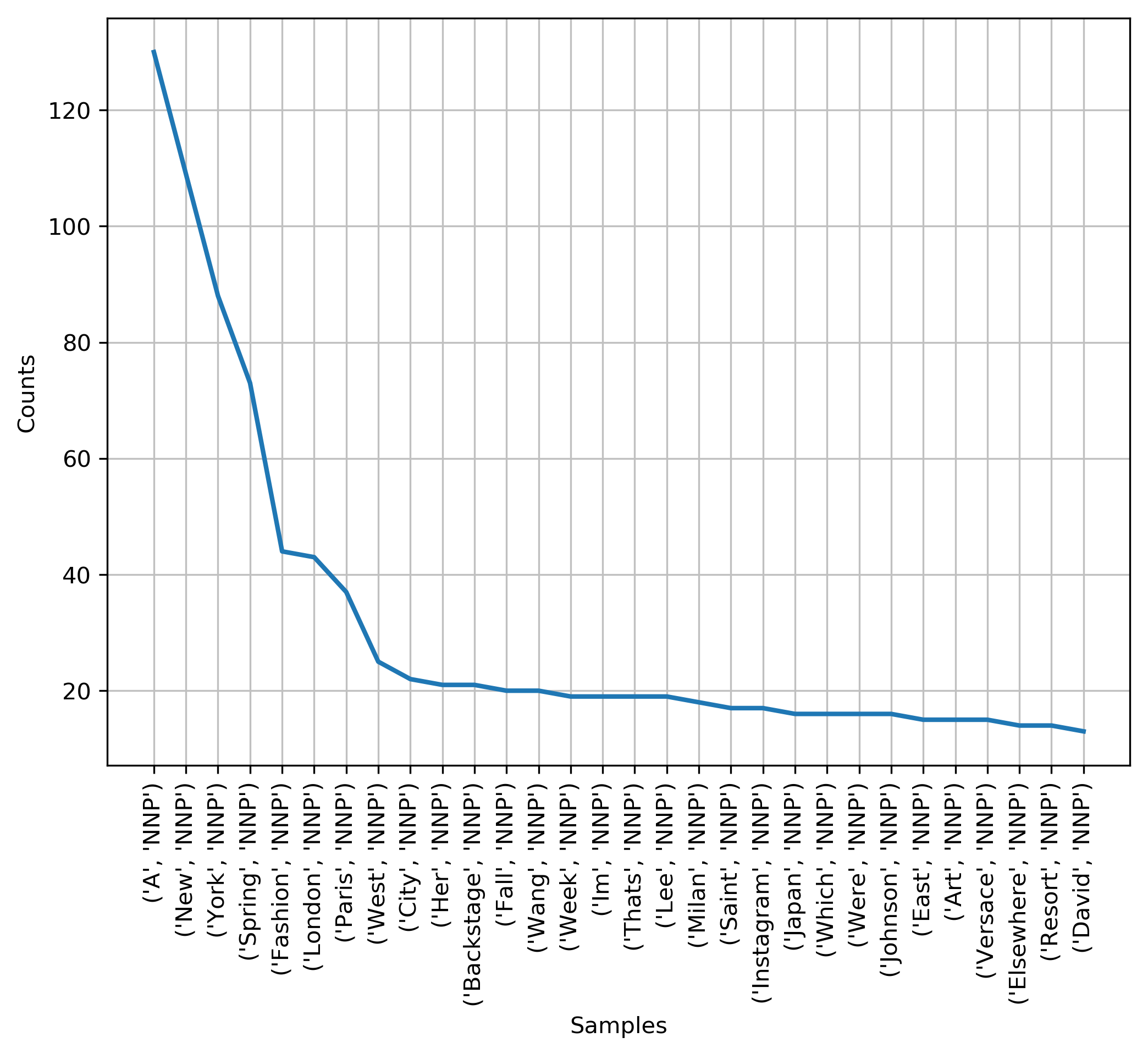
1. POS approach and focus on all the noun forms (NN, NNP, NNS, NNPS)

|  |  |
| --- | --- |
| **Word** | **Frequency** |
| (collection, NN) | 479 |
| (dress, NN) | 267 |
| (dresses, NNS) | 263 |
| (show, NN) | 240 |
| (designer, NN) | 228 |
| (clothes, NNS) | 223 |
| (season, NN) | 219 |
| (way, NN) | 191 |
| (pieces, NNS) | 191 |
| (silk, NN) | 164 |
| (fashion, NN) | 160 |
| (designers, NNS) | 151 |
| (A, NNP) | 130 |
| (time, NN) | 123 |
| (pants, NNS) | 123 |
| (runway, NN) | 119 |
| (Spring, NN) | 119 |
| (looks, NNS) | 118 |
| (skirts, NNS) | 117 |
| (brand, NN) | 110 |
| (New, NNP) | 109 |
| (look, NN) | 109 |
| (prints, NNS) | 108 |
| (women, NNS) | 104 |
| (jacket, NN) | 100 |
| (jackets, NNS) | 100 |
| (point, NN) | 98 |
| (lace, NN) | 97 |
| (collections, NNS) | 96 |
| (leather, NN) | 95 |



1. POS approach and only focus on NNP

|  |  |
| --- | --- |
| **Word** | **Frequency** |
| (A, NNP) | 130 |
| (New, NNP) | 109 |
| (York, NNP) | 88 |
| (Spring, NNP) | 73 |
| (Fashion, NNP) | 44 |
| (London, NNP) | 43 |
| (Paris, NNP) | 37 |
| (West, NNP) | 25 |
| (City, NNP) | 22 |
| (Her, NNP) | 21 |
| (Backstage, NNP) | 21 |
| (Fall, NNP) | 20 |
| (Wang, NNP) | 20 |
| (Week, NNP) | 19 |
| (Im, NNP) | 19 |
| (Thats, NNP) | 19 |
| (Lee, NNP) | 19 |
| (Milan, NNP) | 18 |
| (Saint, NNP) | 17 |
| (Instagram, NNP) | 17 |
| (Japan, NNP) | 16 |
| (Which, NNP) | 16 |
| (Were, NNP) | 16 |
| (Johnson, NNP) | 16 |
| (East, NNP) | 15 |
| (Art, NNP) | 15 |
| (Versace, NNP) | 15 |
| (Elsewhere, NNP) | 14 |
| (Resort, NNP) | 14 |
| (David, NNP) | 13 |



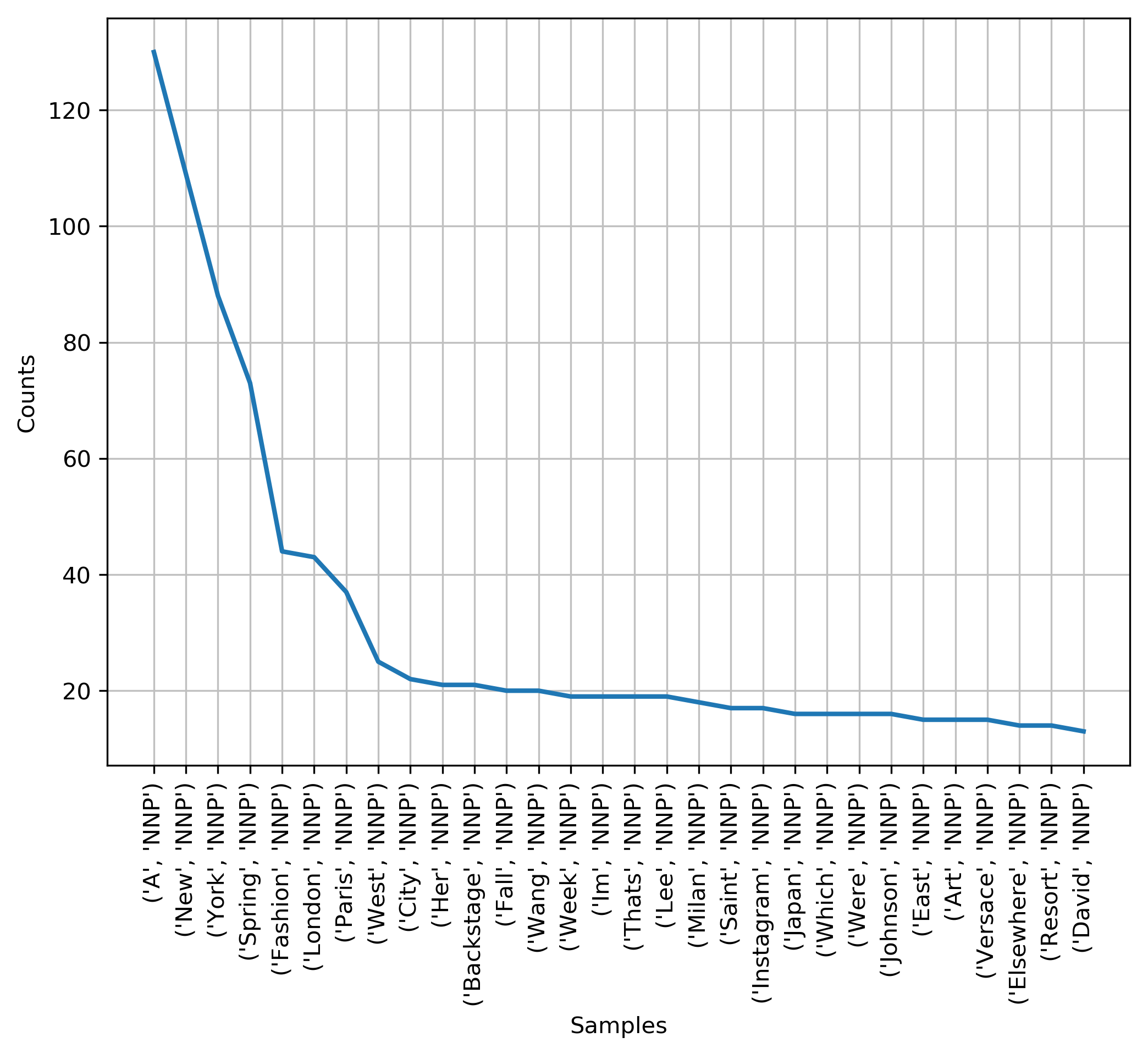
1) Performance of each approach (which one give you best result?)

The **Last one-NNP** gives me the best result and the reason as below:

1. The first top 30 answers contain too many close words like the, and, and of, which are useless for our trend prediction.
2. And the second one, which have been removed stop words and stemmed, also contains many adjectives about the feeling to the dresses in the show and those words could not be a good predictor for our job.
3. For the third one, which outperforms than the first and the second solution. But It contains many general words instead of some specific words that can tell us what is the trend ahead.
4. The fourth is the best solution for us. It gives us the name of those major subjects in the New York show. And those brands which have been lively discussed during the show have a greater exposure to reap the attention of consumers. Thus, more people tend to accept the style of their products and it would be a trend in the near future.

2) What do you see is a fashion trend in 2016. Figures and tables are encouraged in the report.

|  |  |
| --- | --- |
| **Word** | **Frequency** |
| (A, NNP) | 130 |
| (New, NNP) | 109 |
| (York, NNP) | 88 |
| (Spring, NNP) | 73 |
| (Fashion, NNP) | 44 |
| (London, NNP) | 43 |
| (Paris, NNP) | 37 |
| (West, NNP) | 25 |
| (City, NNP) | 22 |
| (Her, NNP) | 21 |
| (Backstage, NNP) | 21 |
| (Fall, NNP) | 20 |
| (Wang, NNP) | 20 |
| (Week, NNP) | 19 |
| (Im, NNP) | 19 |
| (Thats, NNP) | 19 |
| (Lee, NNP) | 19 |
| (Milan, NNP) | 18 |
| (Saint, NNP) | 17 |
| (Instagram, NNP) | 17 |
| (Japan, NNP) | 16 |
| (Which, NNP) | 16 |
| (Were, NNP) | 16 |
| (Johnson, NNP) | 16 |
| (East, NNP) | 15 |
| (Art, NNP) | 15 |
| (Versace, NNP) | 15 |
| (Elsewhere, NNP) | 14 |
| (Resort, NNP) | 14 |
| (David, NNP) | 13 |



From the chart and table above we could easily figure that :

* 1. Firstly, the high frequency of and “Japan”, “Tokyo” ! Triditionally , why “New York”, “London”, “Paris” and “Milan” appear more is that they are big four global show. But the occurence of ‘Japan’ shows that people are watching Japan show in order to get the latest fashion info. If we want to get the fashion trend ahead, we could focus on those brand which would attend the Tokyo show.
  2. Secondly, the high frequent list of designers shows that designer brand’ status is growing , like Alexander Wang, Vera Wang and J.W. Anderson. So people talk about them more would be a trend for thier design works
  3. Third, we can easily what are find the hottest luxury brands in the this New York show, like Chanel, Versace and Gucci. So their products have a bigger chance that more peopl would admire.

3) Brief description of other improvements you can make.

There are two ways we could do better for the job:

1. NER, Named entity recognition (NER) , also known as entity chunking/extraction , is a popular technique used in information extraction to identify and segment the named entities and classify or categorize them under various predefined classes. Using this method, we could extract more insightful knowledges from the Noun words and predict trends well.
2. Sentiment analysis, by getting the sentiment score for each review. We could easily put suitable weight for those important review instead of treating every review fairly regardless the emotional of the critics. Getting insightful knowledge from the most valuable reviews.