

```
In [49]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import os
from statistics import mode, mean
import nltk
#nltk.download('punkt')
#nltk.download('stopwords')
from nltk import sent_tokenize
from nltk import word_tokenize
from nltk.probability import FreqDist
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
file1 = open("C:/Data Science/Assignments/NLP Assignment/novel.txt", "r")
text = file1.read()
print(type(text))
#print(text)
sentences = sent_tokenize(text)
#print(sentences)
words = word_tokenize(text)
print('Total words with out cleaning')
print(len(words))
fdist = FreqDist(words)
#print(fdist.most_common(10))
print('With out cleaning')
fdist.plot(50)
file1.close()
#List of stopwords
stopwords = stopwords.words("english")
#print(stopwords)

#Empty List to store words:
words_no_punc = []
#Removing punctuation marks :
for w in words:
    if w.isalpha():
        words_no_punc.append(w.lower())
#Clean words / Removing Stop Words
clean_words = []
for w in words_no_punc:
```

```

    if w not in stopwords:
        clean_words.append(w)
print('Clean Words')
print(len(clean_words))
#print(clean_words)
fdist = FreqDist(clean_words)
fdist.plot(50)

#Stemming Example :
#Import stemming Library :
porter = PorterStemmer()
#Word-list for stemming :
Stemming_words = []
for w in words:
    Stemming_words.append(porter.stem(w))
print('Length of Stemming words')
print(len(Stemming_words))
print('Stemming_words')
fdist = FreqDist(Stemming_words)
fdist.plot(50)

#Import required Libraries :

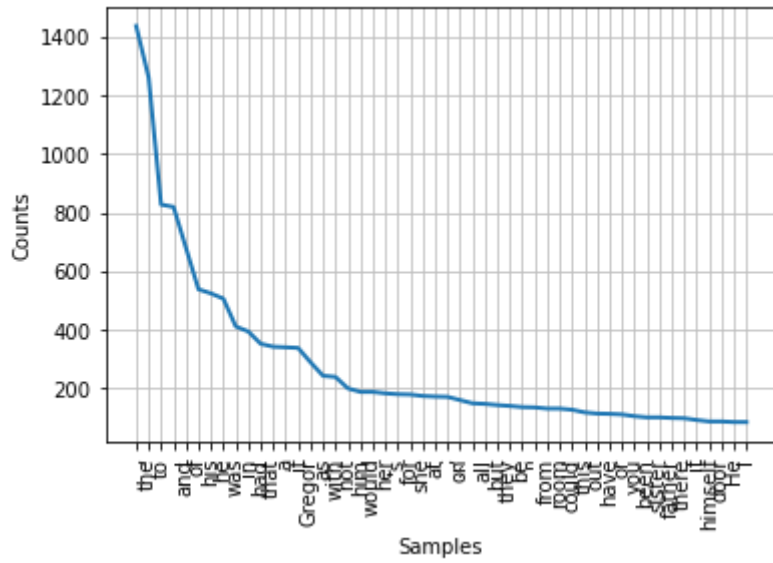
#Sentences for nuemeric vector analysis:
sentences = sent_tokenize(text)
vectorizer = TfidfVectorizer(norm = None)
X= vectorizer.fit_transform(sentences).toarray()
print('Vector vocabulary \n')
print(vectorizer.vocabulary_)
print("\n")
print(vectorizer.get_feature_names_out())
print("\n")
print(X)

```

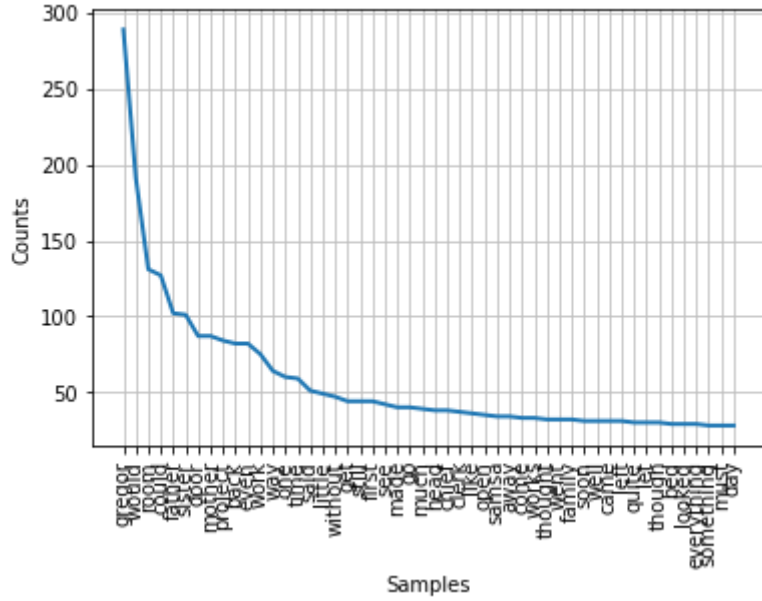
```

<class 'str'>
Total words with out cleaning
28302
With out cleaning

```



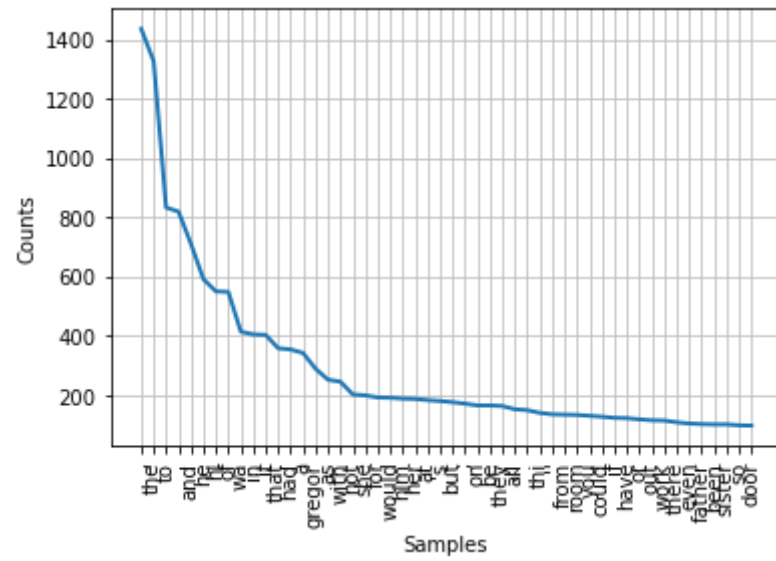
Clean Words
11464



Length of Stemming words

28302

Stemming_words



Vector vocabulary

{'one': 1803, 'morning': 1684, 'when': 2892, 'gregor': 1172, 'samsa': 2226, 'woke': 2937, 'from': 1100, 'troubled': 2716, 'dreams': 796, 'he': 1224, 'found': 1079, 'himself': 1257, 'transformed': 2700, 'in': 1343, 'his': 1259, 'bed': 236, 'into': 1402, 'horrible': 1279, 'vermin': 2811, 'lay': 1484, 'on': 1801, 'armour': 164, 'like': 1531, 'back': 210, 'and': 113, 'if': 1314, 'lifted': 1527, 'head': 1225, 'little': 1549, 'could': 591, 'see': 2259, 'brown': 332, 'belly': 264, 'slightly': 2372, 'domed': 771, 'divided': 762, 'by': 349, 'arches': 155, 'stiff': 2485, 'sections': 2257, 'the': 2610, 'bedding': 237, 'was': 2858, 'hardly': 1210, 'able': 23, 'to': 2665, 'cover': 601, 'it': 1416, 'seemed': 2261, 'ready': 2077, 'slide': 2368, 'off': 1792, 'any': 131, 'moment': 1673, 'many': 1605, 'legs': 1506, 'pitifully': 1926, 'thin': 2621, 'compared': 489, 'with': 2931, 'size': 2356, 'of': 1791, 'rest': 2170, 'him': 1256, 'waved': 2869, 'about': 24, 'helplessly': 1247, 'as': 174, 'looked': 1570, 'what': 2889, 'happened': 1204, 'me': 1615, 'thought': 2632, 'wasn': 2860, 'dream': 795, 'room': 2190, 'proper': 2014, 'human': 1294, 'although': 103, 'too': 2673, 'small': 2377, 'peacefully': 1886, 'between': 274, 'its': 1420, 'four': 1081, 'familiar': 979, 'walls': 2847, 'collection': 472, 'textile': 2604, 'samples': 2225, 'spread': 2441, 'out': 1832, 'table': 2577, 'travelling': 2706, 'salesman': 2219, 'above': 25, 'there': 2615, 'hung': 1300, 'picture': 1920, 'that': 2609, 'had': 1190, 'recently': 2093, 'cut': 629, 'an': 111, 'illustrated': 1320, 'magazine': 1594, 'housed': 1287, 'nice': 1739, 'gilded': 1133, 'frame': 1083, 'showed': 2321, 'lady': 1465, 'fitted': 1025, 'fur': 1109, 'hat': 1218, 'boa': 292, 'who': 2907, 'sat': 2228, 'upright': 2789, 'raising': 2064, 'heavy': 1239, 'muff': 1699, 'covered': 602, 'whole': 2908, 'her': 1249, 'lower': 1586, 'arm': 162, 'towards': 2686, 'viewer': 2818, 'then': 2614, 'turned': 2725, 'look': 1569, 'window': 2921, 'at': 188, 'dull': 811, 'weather': 2879, 'drops': 808, 'rain': 2061, 'be': 226, 'heard': 1231, 'hitting': 1264, 'pane': 1860, 'which': 2898, 'made': 1593, 'feel': 996, 'quite': 2057, 'sad': 2211, 'how': 1290, 'sleep': 2362, 'bit': 279, 'longer': 1567, 'forget': 1064, 'all': 91, 'this': 2628, 'nonsense': 1752, 'but': 345, 'something': 2403, 'unable': 2733, 'do': 764, 'because': 233, 'used': 2796, 'sleeping': 2364, 'right': 2182, 'present': 1980, 'state': 2468, 'couldn': 592, 'get': 1129, 'position': 1955, 'however': 1291, 'hard': 1207, 'threw': 2640, 'onto': 1807, 'always': 105, 'rolled': 2189, 'w here': 2894, 'must': 1703, 'have': 1220, 'tried': 2713, 'hundred': 1298, 'times': 2658, 'shut': 2327, 'eyes': 963, 'so': 2391, 'wouldn': 2958, 'floundering': 1040, 'only': 1806, 'stopped': 2491, 'began': 248, 'mild': 1649, 'pain': 1852, 'never': 1731, 'felt': 1002, 'before': 246, 'oh': 1799, 'god': 1150, 'strenuous': 2502, 'career': 368, 'is': 1414, 've': 2809, 'chosen': 435, 'day': 648, 'doing': 768, 'business': 342, 'takes': 2580, 'much': 1698, 'more': 1682, 'effort': 846, 'than': 2605, 'your': 2978, 'own': 1844, 'home': 1270, 'top': 2677, 'curse': 626, 'worries': 2954, 'making': 1602, 'train': 2691, 'connections': 526, 'bad': 213, 'irregular': 1410, 'food': 1054, 'contact': 542, 'different': 706, 'people': 1894, 'time': 2657, 'you': 2976, 'can': 362, 'know': 1455, 'anyone': 132, 'or': 1818, 'become': 234, 'friendly': 1096, 'them': 2612, 'go': 1148, 'hell': 1242, 'slight': 2370, 'itch': 1417, 'up': 2786, 'pushed': 2043, 'slowly': 2375, 'headboard': 1226, 'lift': 1526, 'better': 273, 'saw': 2235, 'lots': 1578, 'white': 2906, 'spots': 2440, 'didn': 705, 'make': 1600, 'place': 1928, 'drew': 798, 'quickly': 2053, 'soon': 2408, 'touched': 2683, 'overcome': 1838, 'cold': 469, 'shudder': 2325, 'slid': 2367, 'former': 1073, 'getting': 1131, 'early': 823, 'makes': 1601, 'stupid': 2523, 'got': 1158, 'enough': 882, 'other': 1827, 'salesmen': 2220, 'live': 1550, 'life': 1524, 'luxury': 1589, 'for': 1059, 'instance': 1385, 'whenever': 2893, 'guest': 1185, 'house': 1286, 'during': 814, 'copy': 575, 'contract': 553, 'these': 2618, 'gentlemen': 1127, 'are': 158, 'still': 2486, 'sitting': 2352, 'eating': 835, 'their': 2611, 'breakfasts': 318, 'ought': 1829, 'just': 1436, 'try': 2720, 'my': 1705, 'boss': 300, 'kicked': 1443, 'spot': 2439, 'knows': 1458, 'maybe': 1614, 'would': 2957, 'best': 272, 'thing': 2622, 'parents': 1869, 'think': 2624, 'given': 1137, 'notice': 1763, 'long': 1566, 'ago': 78, 'gone': 1154, 'told': 2669, 'tell': 2593, 'everything': 917, 'let': 15

09, 'fall': 976, 'desk': 690, 'funny': 1108, 'sort': 2411, 'talking': 2583, 'down': 784, 'subordinates': 2526, 'especially': 906, 'close': 456, 'hearing': 1233, 'well': 2884, 'some': 2400, 'hope': 1274, 'once': 1802, 'money': 1675, 'together': 2668, 'pay': 1881, 'debt': 654, 'another': 122, 'five': 1026, 'six': 2354, 'years': 2972, 'suppose': 2552, 'definitely': 674, 'll': 1556, 'big': 276, 'change': 402, 'first': 1021, 'though': 2631, 'leaves': 1499, 'over': 1836, 'alarm': 87, 'clock': 455, 'ticking': 2651, 'chest': 425, 'drawers': 793, 'heaven': 1235, 'half': 1194, 'past': 1876, 'hands': 1200, 'were': 2887, 'quietly': 2056, 'moving': 1695, 'forwards': 1078, 'even': 909, 'later': 1478, 'quarter': 2049, 'seven': 2284, 'not': 1759, 'rung': 2203, 'been': 243, 'set': 2283, 'should': 2311, 'certainly': 396, 'yes': 2973, 'possible': 1961, 'through': 2642, 'furniture': 1111, 'rattling': 2068, 'noise': 1747, 'true': 2718, 'slept': 2366, 'probably': 1996, 'deeply': 667, 'now': 1770, 'next': 1738, 'went': 2885, 'catch': 381, 'rush': 2205, 'mad': 1592, 'packed': 1848, 'did': 704, 'particularly': 1872, 'fresh': 1093, 'lively': 1553, 'avoid': 203, 'anger': 1155, 'office': 1796, 'assistant': 182, 'put': 2045, 'report': 2147, 'being': 259, 'man': 1603, 'spineless': 2431, 'no': 1743, 'understanding': 2744, 'reported': 2148, 'sick': 2329, 'extremely': 960, 'strained': 2495, 'suspicious': 2560, 'fifteen': 1006, 'service': 2281, 'yet': 2975, 'ill': 1318, 'come': 475, 'round': 2193, 'doctor': 765, 'medical': 1626, 'insurance': 1389, 'company': 488, 'accuse': 40, 'having': 1222, 'lazy': 1486, 'son': 2407, 'accept': 28, 'recommendation': 2097, 'claim': 441, 'believed': 262, 'ever': 914, 'workshy': 2950, 'entirely': 894, 'wrong': 2965, 'case': 378, 'fact': 968, 'apart': 136, 'excessive': 926, 'sleepiness': 2363, 'after': 69, 'completely': 495, 'hungrier': 1302, 'usual': 2800, 'hurriedly': 1305, 'thinking': 2625, 'decide': 656, 'struck': 2513, 'cautious': 389, 'knock': 1452, 'door': 779, 'near': 1713, 'somebody': 2401, 'called': 353, 'mother': 1686, 'want': 2848, 'somewhere': 2406, 'gentle': 1125, 'voice': 2831, 'shocked': 2303, 'answering': 125, 'recognised': 2095, 'deep': 665, 'inside': 1382, 'painful': 1854, 'uncontrollable': 2739, 'squeaking': 2447, 'mixed': 1667, 'words': 2945, 'echo': 838, 'unclear': 2737, 'leaving': 1500, 'hearer': 1232, 'unsure': 2776, 'whether': 2897, 'properly': 2015, 'wanted': 2849, 'give': 1136, 'full': 1104, 'answer': 123, 'explain': 949, 'circumstances': 439, 'contented': 546, 'saying': 2237, 'thank': 2606, 'noticed': 1766, 'outside': 1834, 'wooden': 2943, 'satisfied': 2230, 'explanation': 952, 'shuffled': 2326, 'away': 206, 'short': 2308, 'conversation': 561, 'members': 1633, 'family': 980, 'aware': 205, 'against': 72, 'expectations': 940, 'father': 985, 'came': 361, 'knocking': 1454, 'side': 2330, 'doors': 780, 'gently': 1128, 'fist': 1022, 'while': 2899, 'again': 71, 'warning': 2854, 'deepness': 668, 'sister': 2348, 'plaintively': 1932, 'aren': 160, 'need': 1722, 'anything': 133, 'answered': 124, 'both': 301, 'sides': 2332, 'remove': 2135, 'strangeness': 2496, 'enunciating': 900, 'very': 2814, 'carefully': 371, 'putting': 2046, 'pauses': 1880, 'each': 820, 'individual': 1366, 'word': 2944, 'breakfast': 317, 'whispered': 2903, 'open': 1808, 'beg': 247, 'opening': 1810, 'instead': 1386, 'congratulated': 524, 'habit': 1189, 'acquired': 45, 'locking': 1563, 'night': 1740, 'peace': 1884, 'without': 2935, 'disturbed': 760, 'dressed': 797, 'most': 1685, 'consider': 531, 'bring': 325, 'thoughts': 2635, 'sensible': 2273, 'conclusions': 513, 'lying': 1590, 'remembered': 2132, 'often': 1798, 'perhaps': 1901, 'caused': 385, 'awkwardly': 208, 'pure': 2039, 'imagination': 1322, 'wondered': 2942, 'imaginings': 1326, 'resolve': 2162, 'themselves': 2613, 'today': 2666, 'slightest': 2371, 'doubt': 783, 'nothing': 1762, 'sign': 2337, 'serious': 2278, 'occupational': 1787, 'hazard': 1223, 'simple': 2342, 'matter': 1610, 'throw': 2644, 'covers': 604, 'blow': 289, 'they': 2619, 'fell': 1001, 'became': 232, 'difficult': 708, 'exceptionally': 925, 'broad': 327, 'arms': 165, 'push': 2042, 'those': 2630, 'continuously': 552, 'directions': 716, 'moreover': 1683, 'control': 559, 'bend': 267, 'stretch': 2503, 'itself': 1421, 'finally': 1012, 'managed': 1604, 'leg': 1503, 'others': 1828, 'free': 1085, 'move': 1691, 'painfully': 1855, 'done': 777, 'said': 2215, 'don': 773, 'keep': 1438, 'trying': 2721, 'part': 1870, 'body': 293, 'seen': 2263, 'imagine': 1323, 'almost': 95, 'frenzy': 1090, 'carelessly': 372, 'shoved': 2319, 'force': 1060, 'gather': 1119, 'chose': 434, 'direction': 715, 'hit': 1263, 'bedpost': 238, 'learned': 1495, 'burning': 340, 'might': 1648, 'sensitive': 2274, 'turning': 2726, 'easily': 831, 'despite': 693, 'breadth': 315, 'weight': 2883, 'bulk': 338, 'eventually': 913, 'followed': 1051, 'last': 1475, 'air':

85, 'occurred': 1790, 'miracle': 1657, 'injured': 1375, 'afraid': 68, 'carry': 375, 'pushing': 2044, 'forward': 1077, 'same': 2224, 'way': 2870, 'price': 1990, 'stay': 2473, 'lose': 1573, 'consciousness': 528, 'took': 2674, 'earlier': 822, 'sighing': 2334, 'watching': 2866, 'struggled': 2515, 'harder': 1208, 'bringing': 326, 'order': 1820, 'chaos': 406, 'whatever': 2890, 'sacrifice': 2209, 'remind': 2133, 'calm': 356, 'consideration': 534, 'rushing': 2207, 'desperate': 692, 'direct': 714, 'clearly': 449, 'unfortunately': 2754, 'narrow': 1710, 'street': 2499, 'enveloped': 901, 'fog': 1048, 'view': 2816, 'confidence': 517, 'cheer': 422, 'offer': 1793, 'already': 98, 'breathing': 322, 'lightly': 1530, 'expected': 941, 'total': 2681, 'stillness': 2487, 'things': 2623, 'real': 2078, 'natural': 1711, 'strikes': 2509, 'will': 2918, 'work': 2946, 'ask': 177, 'task': 2585, 'swinging': 2570, 'entire': 893, 'length': 1507, 'succeeded': 2529, 'falling': 978, 'kept': 1440, 'raised': 2063, 'injuring': 1377, 'happen': 1203, 'carpet': 373, 'main': 1597, 'concern': 509, 'loud': 1579, 'bound': 306, 'raise': 2062, 'risked': 2184, 'sticking': 2483, 'new': 1732, 'method': 1643, 'game': 1116, 'rock': 2186, 'forth': 1075, 'help': 1243, 'two': 2730, 'strong': 2511, 'maid': 1595, 'mind': 1651, 'under': 2741, 'dome': 770, 'peel': 1888, 'load': 1557, 'patient': 1879, 'careful': 370, 'swang': 2563, 'floor': 1039, 'hopefully': 1276, 'find': 1015, 'use': 2795, 'really': 2082, 'call': 352, 'locked': 1562, 'difficulty': 709, 'suppress': 2553, 'smile': 2383, 'moved': 1692, 'far': 982, 'across': 46, 'balance': 215, 'rocked': 2187, 'ten': 2597, 'final': 1011, 'decision': 658, 'ring': 2183, 'flat': 1031, 'someone': 2402, 'froze': 1102, 'danced': 635, 'around': 167, 'remained': 2126, 'quiet': 2054, 're': 2069, 'caught': 383, 'nonsensical': 1753, 'course': 599, 'firm': 1019, 'steps': 2481, 'opened': 1809, 'needed': 1723, 'hear': 1230, 'visitor': 2830, 'greeting': 1171, 'knew': 1449, 'chief': 427, 'clerk': 451, 'why': 2911, 'condemned': 514, 'immediately': 1329, 'highly': 1255, 'shortcoming': 2309, 'employees': 864, 'every': 915, 'louts': 1582, 'faithful': 975, 'devoted': 703, 'pangs': 1861, 'conscience': 527, 'spent': 2427, 'least': 1496, 'couple': 597, 'hours': 1285, 'trainees': 2693, 'enquiries': 883, 'assuming': 184, 'necessarily': 1719, 'show': 2320, 'innocent': 1380, 'trusted': 2719, 'wisdom': 2927, 'investigate': 1407, 'upset': 2790, 'thump': 2647, 'softened': 2394, 'also': 100, 'elastic': 851, 'sound': 2415, 'muffled': 1700, 'noticeable': 1764, 'held': 1241, 'annoyed': 121, 'rubbed': 2198, 'fallen': 977, 'left': 1502, 'concede': 505, 'gruff': 1183, 'reply': 2146, 'question': 2050, 'footsteps': 1058, 'polished': 1952, 'boots': 299, 'adjoining': 60, 'here': 1250, 'daring': 639, 'has': 1217, 'wants': 2851, 'leave': 1498, 'we': 2872, 'say': 2236, 'anyway': 134, 'speak': 2421, 'personally': 1910, 'please': 1942, 'sure': 2554, 'good': 1155, 'forgive': 1066, 'untidiness': 2778, 'mr': 1696, 'isn': 1415, 'continued': 550, 'believe': 261, 'else': 857, 'missed': 1661, 'lad': 1464, 'thinks': 2626, 'nearly': 1717, 'cross': 617, 'goes': 1151, 'evenings': 911, 'town': 2688, 'week': 2881, 'stayed': 2474, 'evening': 910, 'sits': 2351, 'us': 2794, 'kitchen': 1447, 'reads': 2076, 'paper': 1865, 'studies': 2520, 'timetables': 2659, 'idea': 1310, 'relaxation': 2120, 'working': 2948, 'fretsaw': 1094, 'three': 2639, 'amazed': 107, 'hanging': 1202, 'opens': 1812, 'glad': 1139, 'ourselves': 1831, 'stubborn': 2517, 'thoughtfully': 2633, 'miss': 1660, 'explaining': 951, 'mrs': 1697, 'hand': 1198, 'commerce': 482, 'unwell': 2784, 'fortunately': 1076, 'simply': 2343, 'considerations': 535, 'asked': 178, 'impatiently': 1332, 'silence': 2338, 'cry': 621, 'join': 1427, 'she': 2298, 'begun': 253, 'crying': 622, 'danger': 636, 'losing': 1574, 'job': 1425, 'pursue': 2041, 'demands': 680, 'worry': 2955, 'intention': 1394, 'abandoning': 20, 'condition': 515, 'seriously': 2279, 'minor': 1655, 'discourtesy': 731, 'suitable': 2543, 'excuse': 934, 'sacked': 2208, 'disturbing': 761, 'happening': 1205, 'worried': 2953, 'behaviour': 256, 'barricade': 223, 'yourself': 2979, 'causing': 386, 'unnecessary': 2768, 'fail': 969, 'mention': 1636, 'duties': 817, 'unheard': 2756, 'speaking': 2422, 'behalf': 254, 'employer': 865, 'request': 2155, 'clear': 446, 'immediate': 1328, 'am': 106, 'astonished': 187, 'person': 1909, 'suddenly': 2534, 'seem': 2260, 'showing': 2322, 'peculiar': 1887, 'whims': 2900, 'suggest': 2541, 'reason': 2085, 'failure': 971, 'appear': 138, 'entrusted': 899, 'giving': 1138, 'honour': 1273, 'incomprehensible': 1355, 'stubbornness': 2518, 'wish': 2929, 'whatsoever': 2891, 'intercede': 1396, 'nor': 1754, 'secure': 2258, 'originally': 1824, 'intended': 1391, 'private': 1995, 'since': 2345, 'cause': 384, 'waste': 2861, 'learn': 1494, 'turnover': 2727, 'unsatisfactory': 2772, 'lat

e': 1477, 'grant': 1162, 'year': 2970, 'recognise': 2094, 'cannot': 363, 'allow': 92, 'sir': 2347, 'beside': 270, 'forgetting': 1065, 'excitement': 930, 'attack': 191, 'dizziness': 763, 'haven': 1221, 'easy': 832, 'alright': 99, 'shocking': 2304, 'symptom': 2575, 'illness': 1319, 'staying': 2475, 'suffer': 2535, 'basis': 225, 'accusations': 39, 'nobody': 1744, 'read': 2073, 'latest': 1479, 'contracts': 555, 'sent': 2275, 'eight': 848, 'few': 1004, 'strength': 2501, 'wait': 2837, 'recommend': 2096, 'gushed': 1187, 'knowing': 1456, 'practise': 1972, 'insistent': 1384, 'curious': 624, 'sight': 2336, 'responsibility': 2168, 'calmly': 359, 'hurried': 1304, 'station': 2471, 'climb': 453, 'smooth': 2387, 'gave': 1120, 'swing': 2569, 'stood': 2489, 'attention': 194, 'nearby': 1714, 'chair': 398, 'tightly': 2655, 'edges': 840, 'calmed': 357, 'listen': 1544, 'understand': 2743, 'surely': 2555, 'fools': 1056, 'tears': 2591, 'grete': 1174, 'cried': 615, 'communicated': 485, 'straight': 2493, 'quick': 2051, 'spoke': 2436, 'animal': 118, 'calmness': 360, 'contrast': 557, 'screams': 2248, 'anna': 119, 'entrance': 898, 'hall': 1195, 'clapping': 442, 'locksmith': 1564, 'girls': 1135, 'skirts': 2358, 'swishing': 2571, 'ran': 2066, 'wrenching': 2961, 'front': 1101, 'banging': 217, 'homes': 1271, 'awful': 207, 'calmer': 358, 'clearer': 448, 'ears': 827, 'realised': 2081, 'response': 2167, 'situation': 2353, 'confident': 518, 'wise': 2928, 'drawn': 794, 'among': 109, 'great': 1167, 'surprising': 2558, 'achievements': 43, 'distinguish': 751, 'crucial': 620, 'coughed': 588, 'taking': 2581, 'care': 367, 'loudly': 1581, 'coughs': 590, 'judge': 1431, 'meanwhile': 1623, 'whispering': 2904, 'pressed': 1983, 'listening': 1546, 'holding': 1267, 'using': 2799, 'adhesive': 59, 'tips': 2662, 'rested': 2171, 'recover': 2098, 'involved': 1409, 'key': 1441, 'lock': 1561, 'mouth': 1690, 'teeth': 2592, 'grasp': 1164, 'lack': 1462, 'jaw': 1422, 'start': 2461, 'ignoring': 1315, 'kind': 1444, 'damage': 630, 'fluid': 1045, 'flowed': 1042, 'dripped': 801, 'greatly': 1169, 'encouraged': 871, 'calling': 354, 'hold': 1265, 'excitedly': 929, 'following': 1052, 'efforts': 847, 'paying': 1882, 'snapped': 2388, 'break': 316, 'concentration': 507, 'regained': 2111, 'breath': 320, 'handle': 1199, 'wide': 2912, 'turn': 2724, 'double': 782, 'entering': 888, 'occupied': 1788, 'movement': 1693, 'exclaim': 931, 'sounded': 2416, 'soughing': 2413, 'wind': 2920, 'nearest': 1716, 'retreating': 2174, 'driven': 803, 'steady': 2477, 'invisible': 1408, 'hair': 1192, 'dishevelled': 739, 'unfolded': 2752, 'sank': 2227, 'disappeared': 722, 'breast': 319, 'hostile': 1281, 'clenched': 450, 'fists': 1023, 'wanting': 2850, 'uncertainly': 2735, 'living': 1555, 'wept': 2886, 'powerful': 1968, 'shook': 2306, 'leant': 1492, 'bolted': 294, 'along': 97, 'peered': 1889, 'lighter': 1529, 'endless': 875, 'grey': 1175, 'black': 282, 'building': 335, 'hospital': 1280, 'austere': 199, 'regular': 2115, 'line': 1539, 'windows': 2923, 'piercing': 1923, 'facade': 965, 'throwing': 2645, 'large': 1473, 'droplets': 806, 'ground': 1179, 'washing': 2859, 'important': 1336, 'meal': 1616, 'several': 2286, 'reading': 2075, 'number': 1774, 'newspapers': 1737, 'wall': 2845, 'exactly': 920, 'opposite': 1816, 'photograph': 1916, 'lieutenant': 1523, 'army': 166, 'sword': 2572, 'carefree': 369, 'face': 966, 'respect': 2165, 'uniform': 2757, 'bearing': 229, 'landing': 1471, 'stairs': 2452, 'below': 266, 'pack': 1847, 'commercial': 483, 'traveller': 2704, 'arduous': 157, 'earn': 824, 'going': 1152, 'accurately': 38, 'temporarily': 2595, 'remember': 2131, 'achieved': 42, 'removed': 2136, 'diligence': 710, 'our': 1830, 'trapped': 2703, 'take': 2578, 'likes': 1534, 'travellers': 2705, 'enormous': 881, 'wage': 2836, 'soft': 2393, 'prejudice': 1974, 'particular': 1871, 'overview': 1841, 'staff': 2448, 'businessman': 344, 'mistakes': 1665, 'harshly': 1215, 'victim': 2815, 'gossip': 1157, 'chance': 401, 'groundless': 1180, 'complaints': 493, 'impossible': 1340, 'defend': 672, 'usually': 2801, 'arrive': 172, 'exhausted': 937, 'trip': 2714, 'harmful': 1213, 'effects': 845, 'partly': 1873, 'started': 2462, 'protruding': 2022, 'lips': 1543, 'stared': 2459, 'trembling': 2710, 'shoulders': 2313, 'steadily': 2476, 'gradually': 1160, 'secret': 2255, 'prohibition': 2007, 'reached': 2071, 'sudden': 2533, 'foot': 1057, 'rushed': 2206, 'panic': 1862, 'stretched': 2504, 'stairway': 2453, 'supernatural': 2549, 'waiting': 2839, 'save': 2232, 'mood': 1681, 'extreme': 959, 'convinced': 565, 'provide': 2026, 'besides': 271, 'lost': 1576, 'future': 1113, 'won': 2941, 'depended': 682, 'clever': 452, 'lover': 1585, 'women': 2940, 'persuade': 1911, 'talk': 2582, 'considering': 537, 'speech': 2426, 'understood': 2745, 'reach': 2070, 'ridiculously': 2181, 'banister': 218, 'scream': 2245, 'sought': 2414, 'landed': 1470, 'numerous': 1775, 'solid': 239

9, 'pleasure': 1944, 'believing': 263, 'sorrows': 2410, 'end': 872, 'urge': 2792, 'swayed': 2564, 'crouched': 619, 'engrossed': 878, 'herself': 1251, 'jumped': 1433, 'outstretched': 1835, 'fingers': 1017, 'shouting': 2316, 'pity': 1927, 'sake': 2217, 'suggested': 2542, 'unthinking': 2777, 'hurrying': 1307, 'backwards': 212, 'forgotten': 1068, 'behind': 258, 'seeming': 2262, 'coffee': 468, 'pot': 1964, 'knocked': 1453, 'gush': 1186, 'pouring': 1967, 'looking': 1571, 'snapping': 2389, 'jaws': 1423, 'flow': 1041, 'screaming': 2247, 'anew': 114, 'fled': 1033, 'spare': 2420, 'chin': 431, 'run': 2202, 'reaching': 2072, 'leapt': 1493, 'shouts': 2317, 'resounding': 2164, 'staircase': 2451, 'flight': 1037, 'until': 2779, 'relatively': 2119, 'self': 2267, 'controlled': 560, 'running': 2204, 'impeding': 1333, 'seized': 2264, 'stick': 2482, 'overcoat': 1837, 'picked': 1919, 'newspaper': 1736, 'drive': 802, 'stamping': 2455, 'appeals': 137, 'humbly': 1296, 'merely': 1639, 'stamped': 2454, 'chilly': 430, 'pulled': 2035, 'draught': 791, 'flew': 1036, 'curtains': 627, 'fluttered': 1046, 'blown': 290, 'stop': 2490, 'drove': 809, 'hissing': 1262, 'noises': 1748, 'wild': 2916, 'practice': 1971, 'allowed': 93, 'impatient': 1331, 'threat': 2637, 'lethal': 1510, 'choice': 432, 'disgust': 736, 'incapable': 1346, 'frequent': 1091, 'anxious': 129, 'glances': 1142, 'intentions': 1395, 'hinder': 1258, 'tip': 2661, 'distance': 748, 'unbearable': 2734, 'confused': 522, 'finished': 1018, 'mistake': 1664, 'pleased': 1943, 'doorway': 781, 'further': 1112, 'obviously': 1785, 'occur': 1789, 'space': 2419, 'fixed': 1028, 'preparation': 1977, 'pleasant': 1941, 'experience': 947, 'regard': 2112, 'angle': 116, 'flank': 1029, 'scraped': 2242, 'vile': 2822, 'fleck': 1032, 'stuck': 2519, 'fast': 984, 'quivering': 2058, 'hefty': 1240, 'shove': 2318, 'released': 2122, 'flying': 1047, 'heavily': 1237, 'bleeding': 284, 'slammed': 2360, 'ii': 1316, 'dark': 640, 'awoke': 209, 'coma': 473, 'woken': 2938, 'afterwards': 70, 'hadn': 1191, 'fully': 1105, 'impression': 1341, 'leading': 1489, 'light': 1528, 'electric': 854, 'lamps': 1469, 'shone': 2305, 'palely': 1858, 'ceiling': 392, 'tops': 2678, 'feeling': 997, 'clumsily': 462, 'antennae': 126, 'beginning': 252, 'value': 2805, 'scar': 2239, 'limped': 1537, 'badly': 214, 'rows': 2195, 'events': 912, 'dragged': 789, 'lifelessly': 1525, 'actually': 51, 'smell': 2381, 'eat': 833, 'dish': 737, 'filled': 1009, 'sweetened': 2566, 'milk': 1650, 'pieces': 1922, 'bread': 314, 'floating': 1038, 'laughed': 1481, 'dipped': 713, 'covering': 603, 'disappointment': 724, 'tender': 2599, 'worked': 2947, 'snuffling': 2390, 'taste': 2586, 'normally': 1755, 'favourite': 987, 'drink': 800, 'crawled': 609, 'centre': 394, 'crack': 605, 'gas': 1117, 'lit': 1547, 'sometimes': 2404, 'write': 2962, 'recent': 2092, 'lead': 1487, 'gazing': 1121, 'darkness': 642, 'pride': 1991, 'such': 2531, 'wealth': 2876, 'comfort': 477, 'frightening': 1098, 'crawling': 610, 'closed': 457, 'enter': 886, 'waited': 2838, 'resolved': 2163, 'either': 850, 'timorous': 2660, 'vain': 2804, 'previous': 1989, 'everyone': 916, 'unlocked': 2766, 'keys': 1442, 'gaslight': 1118, 'awake': 204, 'distinctly': 750, 'toe': 2667, 'plenty': 1946, 'undisturbed': 2747, 'arrange': 168, 'tall': 2584, 'empty': 868, 'forced': 1061, 'remain': 2125, 'uneasy': 2748, 'shame': 2293, 'couch': 586, 'nonetheless': 1750, 'ease': 829, 'regret': 2114, 'underneath': 2742, 'spent': 2428, 'passed': 1875, 'frequently': 1092, 'hunger': 1301, 'vague': 2803, 'hopes': 1277, 'led': 1501, 'conclusion': 512, 'patience': 1878, 'greatest': 1168, 'bear': 227, 'unpleasantness': 2771, 'impose': 1338, 'opportunity': 1815, 'test': 2603, 'decisions': 659, 'ended': 874, 'anxiously': 130, 'flown': 1044, 'stranger': 2497, 'edge': 839, 'watched': 2865, 'realise': 2080, 'rather': 2067, 'hungry': 1303, 'draw': 792, 'terrible': 2601, 'feet': 1000, 'splashed': 2433, 'surprise': 2557, 'rag': 2059, 'bare': 221, 'carried': 374, 'imagining': 1325, 'wildest': 2917, 'possibilities': 1959, 'guessed': 1184, 'goodness': 1156, 'brought': 331, 'selection': 2266, 'old': 1800, 'rotten': 2192, 'vegetables': 2810, 'bones': 298, 'sauce': 2231, 'raisins': 2065, 'almonds': 94, 'cheese': 424, 'declared': 663, 'inedible': 1367, 'days': 649, 'dry': 810, 'roll': 2188, 'but': 347, 'salt': 2222, 'poured': 1966, 'water': 2867, 'permanently': 1905, 'aside': 176, 'placed': 1929, 'feelings': 998, 'comfortable': 478, 'liked': 1532, 'whirred': 2901, 'injuries': 1376, 'healed': 1228, 'month': 1679, 'finger': 1016, 'knife': 1450, 'hurt': 1308, 'yesterday': 2974, 'less': 1508, 'sucking': 2532, 'greedily': 1170, 'compellingly': 490, 'attracted': 197, 'foods': 1055, 'watering': 2868, 'consumed': 541, 'stand': 2456, 'lethargic': 1511, 'withdrew': 2932, 'startled': 2464, 'asleep': 179, 'rounded': 2194, 'breathe': 321, 'suffocating': 2540, 'bulging': 337,

'unselfconsciously': 2773, 'broom': 329, 'swept': 2568, 'overs': 1840, 'mixing': 1668, 'dropped': 807, 'bin': 277, 'l id': 1520, 'received': 2090, 'second': 2254, 'eaten': 834, 'midday': 1646, 'send': 2269, 'errand': 904, 'starve': 246 7, 'feeding': 995, 'distress': 753, 'indeed': 1358, 'suffering': 2537, 'content': 545, 'sighs': 2335, 'saints': 2216, 'becoming': 235, 'comment': 481, 'construed': 540, 'enjoyed': 880, 'dinner': 712, 'diligently': 711, 'cleared': 447, 'sadly': 2213, 'news': 1734, 'directly': 717, 'rooms': 2191, 'scurry': 2250, 'appropriate': 152, 'press': 1982, 'seld om': 2265, 'mealtime': 1618, 'meals': 1617, 'subject': 2525, 'knees': 1448, 'begged': 249, 'delay': 675, 'within': 29 34, 'hour': 1284, 'tearfully': 2590, 'thanking': 2607, 'dismissal': 741, 'swore': 2573, 'emphatically': 862, 'cookin g': 569, 'bother': 302, 'ate': 189, 'unsuccessfully': 2775, 'receive': 2089, 'thanks': 2608, 'similar': 2341, 'dran k': 790, 'beer': 244, 'hoping': 1278, 'fetch': 1003, 'add': 52, 'selfish': 2268, 'housekeeper': 1289, 'explained': 95 0, 'finances': 1013, 'prospects': 2019, 'receipt': 2088, 'document': 766, 'cash': 379, 'box': 310, 'saved': 2233, 'co llapsed': 470, 'complicated': 497, 'taken': 2579, 'item': 1418, 'incarcerated': 1347, 'misfortune': 1658, 'reduced': 2104, 'despair': 691, 'fiery': 1005, 'vigour': 2821, 'junior': 1434, 'representative': 2151, 'overnight': 1839, 'way s': 2871, 'converted': 564, 'success': 2530, 'benefit': 268, 'delighted': 678, 'splendour': 2434, 'earned': 825, 'cos ts': 584, 'gratitude': 1166, 'warm': 2852, 'affection': 67, 'return': 2175, 'unlike': 2764, 'fond': 1053, 'music': 17 02, 'gifted': 1132, 'expressive': 957, 'violinist': 2826, 'plan': 1933, 'conservatory': 530, 'expense': 945, 'period s': 1903, 'mentioned': 1637, 'lovely': 1584, 'decided': 657, 'planned': 1934, 'grand': 1161, 'announcement': 120, 'ch ristmas': 436, 'totally': 2682, 'pointless': 1950, 'tired': 2663, 'continue': 549, 'wearily': 2877, 'pull': 2034, 'si lent': 2339, 'interrupted': 1401, 'repeated': 2140, 'matters': 1611, 'explanations': 953, 'misfortunes': 1659, 'avail able': 201, 'lot': 1577, 'meantime': 1622, 'interest': 1397, 'accumulated': 36, 'keeping': 1439, 'accumulating': 37, 'nodded': 1745, 'enthusiasm': 891, 'unexpected': 2750, 'thrift': 2641, 'caution': 388, 'surplus': 2556, 'reduce': 210 3, 'freed': 1086, 'closer': 459, 'enable': 870, 'maintain': 1598, 'emergencies': 860, 'healthy': 1229, 'lacking': 146 3, 'holiday': 1269, 'strain': 2494, 'slow': 2374, 'clumsy': 463, 'elderly': 852, 'suffered': 2536, 'asthma': 186, 'st ruggling': 2516, 'sofa': 2392, 'child': 428, 'seventeen': 2285, 'till': 2656, 'enviable': 902, 'consisting': 538, 'we aring': 2878, 'clothes': 461, 'helping': 1245, 'joining': 1429, 'modest': 1670, 'pleasures': 1945, 'playing': 1940, 'violin': 2825, 'cool': 570, 'leather': 1497, 'hot': 1282, 'lie': 1521, 'wink': 2924, 'scratching': 2244, 'climbing': 454, 'sill': 2340, 'propped': 2017, 'leaning': 1491, 'stare': 2458, 'sense': 2271, 'freedom': 1087, 'experienced': 94 8, 'distinct': 749, 'known': 1457, 'lived': 1551, 'charlottenstrasse': 411, 'middle': 1647, 'city': 440, 'barren': 22 2, 'sky': 2359, 'earth': 828, 'mingled': 1654, 'inseparably': 1381, 'observant': 1777, 'twice': 2728, 'exact': 919, 'tidied': 2653, 'inner': 1379, 'easier': 830, 'naturally': 1712, 'pretend': 1985, 'burdensome': 339, 'unpleasant': 27 70, 'entered': 887, 'sooner': 2409, 'precaution': 1973, 'shivering': 2301, 'ordeal': 1819, 'transformation': 2699, 'a ppearance': 139, 'staring': 2460, 'motionless': 1688, 'coming': 480, 'threatened': 2638, 'bite': 280, 'hide': 1253, 'flee': 1034, 'protruded': 2021, 'carrying': 376, 'bedsheet': 242, 'arranged': 169, 'bent': 269, 'sheet': 2299, 'glim psed': 1146, 'arrangement': 170, 'fourteen': 1082, 'appreciated': 147, 'girl': 1134, 'somewhat': 2405, 'useless': 279 7, 'behaved': 255, 'improvement': 1342, 'visit': 2829, 'persuaded': 1912, 'listened': 1545, 'closely': 458, 'approve d': 153, 'unfortunate': 2753, 'courage': 598, 'adult': 62, 'appreciation': 148, 'square': 2445, 'meters': 1642, 'craw l': 608, 'entertain': 889, 'freely': 1088, 'relaxed': 2121, 'happy': 1206, 'letting': 1514, 'crash': 606, 'entertaini ng': 890, 'traces': 2689, 'removing': 2137, 'dare': 637, 'sixteen': 2355, 'bravely': 312, 'cook': 567, 'helped': 124 4, 'unless': 2763, 'choose': 433, 'approached': 151, 'express': 955, 'joy': 1430, 'folds': 1049, 'thrown': 2646, 'ref rained': 2106, 'spying': 2444, 'pair': 1856, 'feeble': 993, 'heaving': 1238, 'heaviest': 1236, 'warnings': 2855, 'las ted': 1476, 'labouring': 1461, 'minutes': 1656, 'saddened': 2212, 'heart': 1234, 'abandoned': 19, 'whose': 2910, 'whe reabouts': 2895, 'tone': 2671, 'added': 53, 'cope': 571, 'comes': 476, 'unchanged': 2736, 'communication': 486, 'mono tonous': 1677, 'months': 1680, 'emptied': 867, 'transform': 2698, 'cave': 390, 'inherited': 1374, 'unimpeded': 2758,

'shaken': 2290, 'influence': 1371, 'mindlessly': 1652, 'loss': 1575, 'advantage': 63, 'agree': 79, 'spokesman': 2438, 'concerned': 510, 'meant': 1621, 'advice': 64, 'sufficient': 2538, 'insist': 1383, 'childish': 429, 'perversity': 1913, 'whereas': 2896, 'age': 73, 'enthusiastic': 892, 'tempted': 2596, 'dominated': 772, 'refused': 2109, 'dissuade': 747, 'writing': 2963, 'groaning': 1177, 'poked': 1951, 'considerate': 533, 'pulling': 2036, 'inch': 1348, 'startlement': 2465, 'prevent': 1987, 'attract': 196, 'assure': 185, 'unusual': 2782, 'admit': 61, 'fro': 1099, 'calls': 355, 'scraping': 2243, 'assailed': 180, 'emptying': 869, 'dear': 652, 'containing': 544, 'tools': 2675, 'worn': 2952, 'homework': 1272, 'trainee': 2692, 'high': 1254, 'school': 2241, 'infant': 1368, 'stepped': 2480, 'catching': 382, 'sallied': 2221, 'changed': 403, 'denuded': 681, 'copious': 574, 'glass': 1144, 'firmly': 1020, 'watch': 2864, 'shall': 2292, 'met': 1640, 'albeit': 89, 'tremor': 2712, 'safe': 2214, 'chase': 413, 'unyielding': 2785, 'jump': 1432, 'patch': 1877, 'flowers': 1043, 'wallpaper': 2846, 'screamed': 2246, 'immobile': 1330, 'shouted': 2315, 'glowering': 1147, 'shaking': 2291, 'spoken': 2437, 'smelling': 2382, 'salts': 2223, 'faint': 972, 'advise': 65, 'various': 2808, 'bottles': 304, 'bottle': 303, 'broke': 328, 'splinter': 2435, 'caustic': 387, 'medicine': 1627, 'delaying': 676, 'death': 653, 'oppressed': 1817, 'anxiety': 128, 'reproach': 2152, 'confusion': 523, 'spin': 2429, 'numb': 1773, 'arrived': 173, 'subdued': 2524, 'openly': 1811, 'fainted': 973, 'mean': 1619, 'responsible': 2169, 'act': 47, 'violence': 2824, 'disappear': 721, 'subtleties': 2528, 'ah': 82, 'sounding': 2417, 'angry': 117, 'imagined': 1324, 'neglected': 1726, 'laying': 1485, 'entombed': 897, 'trips': 2715, 'armchair': 163, 'nightgown': 1742, 'walk': 2841, 'sunday': 2544, 'public': 2033, 'wrapped': 2960, 'labour': 1460, 'walking': 2843, 'invariably': 1406, 'companions': 487, 'standing': 2457, 'smart': 2379, 'blue': 291, 'gold': 1153, 'buttons': 348, 'banking': 220, 'institute': 1387, 'collar': 471, 'coat': 465, 'emerged': 859, 'bushy': 341, 'eyebrows': 961, 'alert': 90, 'unkempt': 2761, 'combed': 474, 'scalp': 2238, 'cap': 364, 'monogram': 1676, 'bank': 219, 'arc': 154, 'trouser': 2717, 'pockets': 1947, 'bottom': 305, 'determination': 700, 'walked': 2842, 'unusually': 2783, 'soles': 2396, 'wasted': 2862, 'strict': 2506, 'scurried': 2249, 'decisive': 660, 'largely': 1474, 'feared': 989, 'provoking': 2032, 'step': 2479, 'countless': 594, 'movements': 1694, 'noticeably': 1765, 'lungs': 1587, 'reliable': 2123, 'lurched': 1588, 'muster': 1704, 'saving': 2234, 'forgot': 1067, 'concealed': 504, 'carved': 377, 'notches': 1760, 'protrusions': 2023, 'tossed': 2680, 'apple': 143, 'shock': 2302, 'point': 1948, 'bombard': 295, 'fruit': 1103, 'bowl': 309, 'sideboard': 2331, 'aim': 84, 'red': 2099, 'apples': 144, 'motors': 1689, 'glanced': 1141, 'harm': 1211, 'squarely': 2446, 'lodged': 1565, 'drag': 788, 'incredible': 1357, 'changing': 405, 'nailed': 1706, 'senses': 2272, 'blouse': 288, 'unfastened': 2751, 'sliding': 2369, 'stumbling': 2522, 'uniting': 2760, 'ability': 22, 'begging': 250, 'iii': 1317, 'dared': 638, 'flesh': 1035, 'visible': 2828, 'reminder': 2134, 'injury': 1378, 'current': 625, 'revolting': 2178, 'form': 1070, 'member': 1632, 'treated': 2708, 'enemy': 877, 'contrary': 556, 'duty': 818, 'swallow': 2561, 'revulsion': 2179, 'mobility': 1669, 'ancient': 112, 'invalid': 1404, 'deterioration': 699, 'opinion': 1813, 'permission': 1906, 'thus': 2650, 'differently': 707, 'conversations': 562, 'ones': 1804, 'longing': 1568, 'damp': 633, 'hotel': 1283, 'nowadays': 1771, 'lamp': 1468, 'sew': 2287, 'fancy': 981, 'underwear': 2746, 'fashion': 983, 'shop': 2307, 'sales': 2218, 'shorthand': 2310, 'french': 1089, 'wake': 2840, 'sewing': 2288, 'dozing': 786, 'exchange': 927, 'grin': 1176, 'unused': 2781, 'peg': 1892, 'slumber': 2376, 'serve': 2280, 'expecting': 942, 'superior': 2548, 'result': 2173, 'shabbier': 2289, 'stains': 2450, 'shiny': 2300, 'uncomfortable': 2738, 'peaceful': 1885, 'obstinate': 1781, 'regularly': 2116, 'importune': 1337, 'reproaches': 2153, 'refusing': 2110, 'tug': 2722, 'sleeve': 2365, 'whisper': 2902, 'endearments': 873, 'ear': 821, 'effect': 844, 'sink': 2346, 'deeper': 666, 'abruptly': 26, 'supported': 2551, 'needle': 1724, 'pen': 1893, 'overworked': 1842, 'absolutely': 27, 'household': 1288, 'budget': 334, 'smaller': 2378, 'dismissed': 742, 'thick': 2620, 'boned': 297, 'charwoman': 412, 'flapped': 1030, 'amount': 110, 'hoped': 1275, 'items': 1419, 'jewellery': 1424, 'belonging': 265, 'sold': 2395, 'functions': 1106, 'celebrations': 393, 'loudest': 1580, 'complaint': 492, 'imaginable': 1321, 'transferring': 2697, 'addresses': 57, 'reasons': 2087, 'transport': 2702, 'crate': 607, 'holes': 1268, 'related': 2118, 'world': 2951, 'expects': 9

43, 'poor': 1954, 'sacrificed': 2210, 'strangers': 2498, 'behest': 257, 'customers': 628, 'sit': 2349, 'cheek': 420, 'mingle': 1653, 'eyed': 962, 'affairs': 66, 'apprentices': 149, 'teaboy': 2589, 'friends': 1097, 'businesses': 343, 'chambermaids': 400, 'provincial': 2029, 'memory': 1634, 'appeared': 140, 'cashier': 380, 'whom': 2909, 'inaccessibl e': 1344, 'rage': 2060, 'shown': 2323, 'plans': 1935, 'pantry': 1864, 'entitled': 895, 'sweep': 2565, 'indifferent': 1363, 'untouched': 2780, 'quicker': 2052, 'smears': 2380, 'dirt': 719, 'balls': 216, 'dust': 815, 'filth': 1010, 'wor st': 2956, 'places': 1930, 'weeks': 2882, 'touchy': 2684, 'cleaning': 445, 'alone': 96, 'thoroughly': 2629, 'clean': 443, 'bucketfuls': 333, 'dampness': 634, 'bitter': 281, 'punished': 2037, 'aggrieved': 76, 'mothers': 1687, 'implorin g': 1335, 'convulsive': 566, 'helpless': 1246, 'agitated': 77, 'accused': 41, 'bedroom': 239, 'quaking': 2047, 'thump ed': 2648, 'hissed': 1261, 'closing': 460, 'widow': 2914, 'robust': 2185, 'bone': 296, 'structure': 2514, 'withstan d': 2936, 'hardest': 1209, 'repelled': 2142, 'curiosity': 623, 'chasing': 414, 'amazement': 108, 'crossed': 618, 'fai led': 970, 'briefly': 324, 'considered': 536, 'dung': 813, 'beetle': 245, 'responded': 2166, 'disturb': 759, 'windowp anes': 2922, 'indicating': 1362, 'spring': 2442, 'resentful': 2160, 'toward': 2685, 'infirm': 1369, 'chairs': 399, 'i ntending': 1392, 'corner': 579, 'prepared': 1979, 'play': 1937, 'spit': 2432, 'changes': 404, 'anywhere': 135, 'rente d': 2139, 'earnest': 826, 'beards': 228, 'peering': 1890, 'tidy': 2654, 'establishment': 908, 'clutter': 464, 'tolera te': 2670, 'dirty': 720, 'furnishings': 1110, 'equipment': 903, 'superfluous': 2547, 'discard': 725, 'dustbins': 816, 'hurry': 1306, 'chuck': 437, 'object': 1776, 'woman': 2939, 'likely': 1533, 'junk': 1435, 'enjoy': 879, 'lain': 1466, 'darkest': 641, 'formerly': 1074, 'serviettes': 2282, 'knives': 1451, 'forks': 1069, 'meat': 1624, 'piled': 1924, 'po tatoes': 1965, 'steaming': 2478, 'dishes': 738, 'gentleman': 1126, 'count': 593, 'authority': 200, 'piece': 1921, 'wi shing': 2930, 'establish': 907, 'sufficiently': 2539, 'cooked': 568, 'satisfaction': 2229, 'smiled': 2384, 'bowed': 3 08, 'mumbled': 1701, 'perfect': 1895, 'remarkable': 2129, 'chewing': 426, 'perform': 1896, 'toothless': 2676, 'feed': 994, 'dying': 819, 'throughout': 2643, 'played': 1938, 'produced': 2001, 'page': 1849, 'smoking': 2386, 'attentive': 195, 'hallway': 1196, 'young': 2977, 'cosy': 585, 'love': 1583, 'player': 1939, 'begin': 251, 'therefore': 2617, 'exa ggerated': 921, 'courtesy': 600, 'offered': 1794, 'seat': 2252, 'paid': 1851, 'thoughtless': 2634, 'hidden': 1252, 'e verywhere': 918, 'threads': 2636, 'hairs': 1193, 'remains': 2128, 'wipe': 2925, 'shy': 2328, 'immaculate': 1327, 'pre occupied': 1976, 'notes': 1761, 'withdrew': 2933, 'heads': 1227, 'sunk': 2545, 'volume': 2833, 'observed': 1778, 'obv ious': 1784, 'beautiful': 230, 'disappointed': 723, 'performance': 1897, 'politeness': 1953, 'unnerving': 2769, 'ble w': 285, 'smoke': 2385, 'cigarettes': 438, 'upwards': 2791, 'noses': 1757, 'beautifully': 231, 'lines': 1540, 'melanc holy': 1631, 'expression': 956, 'meet': 1629, 'captivate': 365, 'unknown': 2762, 'nourishment': 1769, 'yearning': 297 1, 'determined': 702, 'skirt': 2357, 'hiss': 1260, 'attackers': 192, 'refuse': 2108, 'emotion': 861, 'shoulder': 231 2, 'kiss': 1445, 'neck': 1720, 'necklace': 1721, 'pointing': 1949, 'wasting': 2863, 'forefinger': 1062, 'driving': 80 4, 'attempted': 193, 'block': 286, 'dawning': 647, 'realisation': 2079, 'neighbour': 1728, 'tugged': 2723, 'drop': 80 5, 'bow': 307, 'hang': 1201, 'limply': 1538, 'instrument': 1388, 'lap': 1472, 'laboriously': 1459, 'pressure': 1984, 'pillows': 1925, 'beds': 241, 'slipped': 2373, 'obsessed': 1779, 'owed': 1843, 'tenants': 2598, 'urged': 2793, 'thund er': 2649, 'thereby': 2616, 'halt': 1197, 'declare': 662, 'glancing': 1143, 'gain': 1114, 'repugnant': 2154, 'conditi ons': 516, 'prevail': 1986, 'decisively': 661, 'proceed': 1998, 'action': 48, 'damages': 632, 'grounds': 1181, 'ahea d': 83, 'joined': 1428, 'staggered': 2449, 'stretching': 2505, 'nap': 1709, 'uncontrolled': 2740, 'nodding': 1746, 'w eak': 2873, 'introduction': 1403, 'monster': 1678, 'brother': 330, 'rid': 2180, 'humanly': 1295, 'cough': 587, 'dull y': 812, 'deranged': 684, 'forehead': 1063, 'definite': 673, 'ideas': 1311, 'plates': 1936, 'occasionally': 1786, 'co ughing': 589, 'tortured': 2679, 'endure': 876, 'wiped': 2926, 'mechanical': 1625, 'sympathy': 2574, 'shrugged': 2324, 'helplessness': 1248, 'displacing': 743, 'certainty': 397, 'vigorously': 2820, 'acceptance': 29, 'harmed': 1212, 'bei ngs': 260, 'lives': 1554, 'persecuting': 1908, 'streets': 2500, 'starting': 2463, 'beyond': 275, 'comprehension': 50 0, 'willing': 2919, 'excited': 928, 'protect': 2020, 'startling': 2466, 'wracked': 2959, 'required': 2157, 'deal': 65

1, 'repeatedly': 2141, 'striking': 2510, 'alarmed': 88, 'unhappy': 2755, 'exhaustion': 938, 'panting': 1863, 'separated': 2277, 'noticing': 1767, 'concentrated': 506, 'distract': 752, 'glance': 1140, 'sprung': 2443, 'discovery': 734, 'spindly': 2430, 'unnatural': 2767, 'aching': 44, 'weaker': 2874, 'altogether': 104, 'decayed': 655, 'inflamed': 137, 'area': 159, 'strongly': 2512, 'rumination': 2201, 'tower': 2687, 'strike': 2508, 'weakly': 2875, 'nostrils': 175, 'cleaner': 444, 'slamming': 2361, 'brief': 323, 'special': 2423, 'purpose': 2040, 'martyr': 1609, 'attributed': 198, 'tickle': 2652, 'nuisance': 1772, 'resistance': 2161, 'whistled': 2905, 'yank': 2969, 'shout': 2314, 'bedrooms': 240, 'ave': 202, 'dead': 650, 'stone': 2488, 'marriage': 1608, 'blanket': 283, 'nightdress': 1741, 'paleness': 1859, 'confirm': 519, 'enquiringly': 884, 'checked': 417, 'checking': 418, 'replied': 2145, 'prove': 2025, 'sending': 2270, 'sideways': 2333, 'complete': 494, 'example': 923, 'corpse': 581, 'dried': 799, 'pained': 1853, 'warmth': 2853, 'march': 1606, 'irritably': 1411, 'men': 1635, 'coats': 466, 'wife': 2915, 'daughter': 645, 'disconcerted': 729, 'sweetly': 2567, 'backs': 211, 'continually': 548, 'gleeful': 1145, 'anticipation': 127, 'quarrel': 2048, 'favour': 986, 'contents': 547, 'rearranging': 2084, 'positions': 1956, 'humility': 1297, 'strides': 2507, 'rubbing': 2199, 'friend': 1095, 'fear': 988, 'connection': 525, 'leader': 1488, 'hats': 1219, 'sticks': 2484, 'holder': 1266, 'premises': 1975, 'mistrust': 1666, 'leaned': 1490, 'progress': 2006, 'reappear': 2083, 'moments': 1674, 'butcher': 346, 'boy': 311, 'proud': 2024, 'posture': 1963, 'tray': 2707, 'nearer': 1715, 'relieved': 2124, 'wrote': 2966, 'letters': 1513, 'excuse': 933, 'employers': 866, 'contractor': 554, 'principal': 1992, 'irritation': 1412, 'tremendous': 2711, 'vertical': 2813, 'ostrich': 1826, 'feather': 990, 'source': 2418, 'laugh': 1480, 'needn': 1725, 'sorted': 2412, 'intent': 1393, 'continuing': 551, 'describing': 689, 'detail': 698, 'prevented': 1988, 'telling': 2594, 'peevied': 1891, 'cheerio': 423, 'sharply': 2297, 'terribly': 2602, 'tonight': 2672, 'gets': 1130, 'destroyed': 696, 'gained': 1115, 'twisted': 2729, 'stuff': 2521, 'kissed': 1446, 'hugged': 1293, 'tram': 2694, 'country': 596, 'sunshine': 2546, 'comfortably': 479, 'seats': 2253, 'discussed': 735, 'examination': 922, 'jobs': 1426, 'promise': 2010, 'cheaper': 415, 'location': 1559, 'practical': 1969, 'livelier': 1552, 'cheeks': 421, 'pale': 1857, 'simultaneously': 2344, 'blossoming': 287, 'built': 336, 'quieter': 2055, 'agreed': 80, 'confirmation': 520, 'destination': 694, 'project': 2008, 'gutenberg': 1188, 'ebook': 836, 'metamorphosis': 1641, 'franz': 1084, 'kafka': 1437, 'translated': 2701, 'david': 646, 'wyllie': 2968, 'file': 1007, 'named': 1708, '5200': 9, 'txt': 2731, 'zip': 2980, 'associated': 183, 'files': 1008, 'formats': 1072, 'http': 1292, 'www': 2967, 'net': 1729, 'updated': 2787, 'editions': 842, 'replace': 2143, 'renamed': 2138, 'creating': 612, 'works': 2949, 'domain': 769, 'print': 1993, 'means': 1620, 'owns': 1846, 'united': 2759, 'states': 2470, 'copyright': 577, 'foundation': 1080, 'distribute': 754, 'royalties': 2196, 'rules': 2200, 'general': 1123, 'terms': 2600, 'license': 1518, 'apply': 146, 'copying': 576, 'distributing': 756, 'tm': 2664, 'electronic': 855, 'concept': 508, 'trademark': 2690, 'registered': 2113, 'may': 1613, 'charge': 407, 'ebooks': 837, 'specific': 2424, 'copies': 573, 'complying': 499, 'creation': 613, 'derivative': 685, 'reports': 2149, 'performances': 1898, 'research': 2159, 'modified': 1672, 'printed': 1994, 'practically': 1970, 'redistribution': 2102, 'mission': 1662, 'promoting': 2011, 'distribution': 757, 'phrase': 1917, 'comply': 498, 'online': 1805, 'section': 2256, 'redistributing': 2101, 'indicate': 1361, 'intellectual': 1390, 'property': 2016, 'agreement': 81, 'abide': 21, 'cease': 391, 'destroy': 695, 'possession': 1958, 'fee': 992, 'obtaining': 1783, 'access': 32, 'obtain': 1782, 'refund': 2107, 'entity': 896, 'paragraph': 1867, 'follow': 1050, 'preserve': 1981, 'literary': 1548, 'archive': 156, 'pglaf': 1915, 'compilation': 491, 'located': 1558, 'performing': 1900, 'displaying': 746, 'based': 224, 'references': 2105, 'support': 2550, 'sharing': 2296, 'compliance': 496, 'name': 1707, 'format': 1071, 'attached': 190, 'share': 2294, 'copyrighted': 578, 'included': 1351, 'information': 1372, 'owner': 1845, 'imposed': 1339, 'laws': 1483, 'govern': 1159, 'countries': 595, 'constant': 539, 'check': 416, 'addition': 54, 'downloading': 785, 'representations': 2150, 'concerning': 511, 'status': 2472, 'sentence': 2276, 'active': 49, 'links': 1542, 'prominently': 2009, 'appears': 142, 'accessed': 33, 'displayed': 745, 'performed': 1899, 'viewed': 2817, 'copied': 572, 'distributed': 755, 'cost': 583, 'restrictions': 2172, 'derived': 687, 'doe

s': 767, 'contain': 543, 'posted': 1962, 'fees': 999, 'charges': 408, 'providing': 2028, 'appearing': 141, 'requirements': 2158, 'paragraphs': 1868, 'additional': 55, 'linked': 1541, 'unlink': 2765, 'detach': 697, 'display': 744, 'redistribute': 2100, 'convert': 563, 'binary': 278, 'compressed': 501, 'marked': 1607, 'nonproprietary': 1751, 'proprietary': 2018, 'including': 1353, 'processing': 1999, 'hypertext': 1309, 'plain': 1931, 'vanilla': 2806, 'ascii': 175, 'official': 1797, 'version': 2812, 'web': 2880, 'site': 2350, 'user': 2798, 'exporting': 954, 'upon': 2788, 'original': 1823, 'alternate': 102, 'include': 1350, 'specified': 2425, 'viewing': 2819, 'reasonable': 2086, 'provided': 2027, 'royalty': 2197, '20': 3, 'gross': 1178, 'profits': 2005, 'derive': 686, 'calculated': 351, 'calculate': 350, 'applicable': 145, 'taxes': 2588, 'donate': 774, 'payments': 1883, '60': 11, 'date': 644, 'prepare': 1978, 'legally': 1505, 'periodic': 1902, 'tax': 2587, 'returns': 2176, 'donations': 776, 'notifies': 1768, 'mail': 1596, '30': 5, 'require': 2156, 'possessed': 1957, 'physical': 1918, 'medium': 1628, 'discontinue': 730, 'accordance': 35, 'replacement': 2144, 'defect': 669, 'discovered': 733, '90': 17, 'group': 1182, 'michael': 1645, 'hart': 1216, 'volunteers': 2835, 'expend': 944, 'considerable': 532, 'identify': 1313, 'transcribe': 2695, 'proofread': 2013, 'stored': 2492, 'defects': 671, 'limited': 1536, 'incomplete': 1354, 'inaccurate': 1345, 'corrupt': 582, 'data': 643, 'transcription': 2696, 'errors': 905, 'infringement': 1373, 'defective': 670, 'damaged': 631, 'disk': 740, 'computer': 502, 'virus': 2827, 'codes': 467, 'warranty': 2857, 'disclaimer': 727, 'except': 924, 'described': 688, 'party': 1874, 'disclaim': 726, 'liability': 1515, 'expenses': 946, 'legal': 1504, 'remedies': 2130, 'negligence': 1727, 'breach': 313, 'f3': 964, 'distributor': 758, 'liable': 1516, 'actual': 50, 'indirect': 1364, 'consequential': 529, 'punitive': 2038, 'incidental': 1349, 'possibility': 1960, 'discover': 732, 'receiving': 2091, 'written': 2964, 'elect': 853, 'lieu': 1522, 'electronically': 856, 'demand': 679, 'opportunities': 1814, 'fix': 1027, 'problem': 1997, 'warranties': 2856, 'implied': 1334, 'merchantability': 1638, 'fitness': 1024, 'disclaimers': 728, 'certain': 395, 'exclusion': 932, 'limitation': 1535, 'types': 2732, 'violates': 2823, 'law': 1482, 'interpreted': 1400, 'maximum': 1612, 'permitted': 1907, 'invalidity': 1405, 'unenforceability': 2749, 'provision': 2030, 'void': 2832, 'remaining': 2127, 'provisions': 2031, 'indemnity': 1360, 'indemnify': 1359, 'agent': 75, 'employee': 863, 'production': 2002, 'promotion': 2012, 'harmless': 1214, 'arise': 161, 'indirectly': 1365, 'alteration': 101, 'modification': 1671, 'additions': 56, 'deletions': 677, 'synonymous': 2576, 'readable': 2074, 'widest': 2913, 'variety': 2807, 'computers': 503, 'obsolete': 1780, 'aged': 74, 'exists': 939, 'hundreds': 1299, 'walks': 2844, 'financial': 1014, 'assistance': 181, 'critical': 616, 'goals': 1149, 'ensuring': 885, 'generations': 1124, '2001': 4, 'created': 611, 'permanent': 1904, 'org': 1821, 'non': 1749, 'profit': 2004, '501': 8, 'educational': 843, 'corporation': 580, 'organized': 1822, 'mississippi': 1663, 'granted': 1163, 'exempt': 936, 'internal': 1398, 'revenue': 2177, 'ein': 849, 'federal': 991, 'identification': 1312, '64': 13, '6221541': 12, 'letter': 1512, 'fundraising': 1107, 'contributions': 558, 'deductible': 664, 'extent': 958, '4557': 6, 'melan': 1630, 'dr': 787, 'fairbanks': 974, 'ak': 86, '99712': 18, 'scattered': 2240, 'locations': 1560, '809': 15, 'north': 1756, '1500': 1, 'west': 2888, 'lake': 1467, 'ut': 2802, '84116': 16, '801': 14, '596': 10, '1887': 2, 'email': 858, 'gregory': 1173, 'newby': 1733, 'executive': 935, 'director': 718, 'gbnewby': 1122, 'depends': 683, 'survive': 2559, 'increasing': 1356, 'licensed': 1519, 'machine': 1591, 'accessible': 34, 'array': 171, 'outdated': 1833, '000': 0, 'maintaining': 1599, 'irs': 1413, 'committed': 484, 'regulating': 2117, 'charities': 410, 'charitable': 409, '50': 7, 'paperwork': 1866, 'solicit': 2397, 'determine': 701, 'solicitation': 2398, 'accepting': 31, 'unsolicited': 2774, 'donors': 778, 'approach': 150, 'offers': 1795, 'international': 1399, 'gratefully': 1165, 'accepted': 30, 'statements': 2469, 'treatment': 2709, 'swamp': 2562, 'pages': 1850, 'donation': 775, 'methods': 1644, 'addresses': 58, 'checks': 419, 'credit': 614, 'card': 366, 'professor': 2003, 'originator': 1825, 'library': 1517, 'shared': 2295, 'thirty': 2627, 'loose': 1572, 'network': 1730, 'volunteer': 2834, 'confirmed': 521, 'necessarily': 1718, 'edition': 841, 'pg': 1914, 'search': 2251, 'facility': 967, 'includes': 1352, 'produce': 2000, 'subscribe': 2527, 'newsletter': 1735}

```
['000' '1500' '1887' ... 'your' 'yourself' 'zip']
```

```
[[0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]  
 ...  
 [0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]]
```

In []:

In []: