

# Problem Statement

Vinit Patel

February 26, 2021

# 1 Problem Statement

Since the explosion of the internet, the ability to communicate more efficiently between people has increased exponentially. This has caused language to develop at a much more rapid pace with new words and new meanings to already existing words being created all across the world. Looking at just English, in the figure below, it is apparent that since the 12th century the language has gained just below 100,000 new words but only lost 2,500 words [2]. This is caused by the fact that three-quarters of the language is borrowed from classical or romance languages and the last quarter is from 350 other languages [1, pg.23]. Now combined with access to a global network where one person can interact with thousands and millions, the everyday average person has the same ability to create new words or change the meaning of words and popularize it over night. This was once only able to be done by influential figures and authors, an example would be Shakespeare who is believed to have created 1,700 words himself. Now it does not matter whether they speak the language or not, since the globalization of languages is causing overlap over all the languages. In the same breath words can start to lose meaning overnight and can sometimes be forgotten forever very quickly. Saving these flickers of the lexicon allows history to be told in a more comprehensive manner with greater ability to understand causes of these flickers and the evolution of language across regions of the world.

Period	First Use		Last Use		In Use	
	Whole Words	Any Sense	Whole Words	Any Sense	Whole Words	Any Sense
-1149 (Old English)	4308	3836	0	139	4308	3836
1150-1349 (Middle English)	6903	8718	0	0	11,211	10,714
1350-1499 (Late Middle English)	11,755	16,666	23	1539	22,966	22,484
1470-1499	1798	4052	8	524	24,741	23,826
1500-1529	1624	3906	4	435	26,357	25,366
1530-1569	4880	9877	29	1117	31,233	30,283
1570-1599	6323	13,528	173	2041	37,527	36,542
1600-1629	5916	12,720	168	2951	43,270	42,185
1630-1669	5498	11,861	151	3827	48,600	47,340
1670-1699	3042	7071	79	2793	51,491	49,883
1700-1729	2378	6065	402	2833	53,790	52,035
1730-1769	2724	6594	403	2485	56,112	54,246
1770-1799	3520	7632	279	1836	59,229	57,329
1800-1829	5181	10,940	323	1832	64,131	62,372
1830-1869	10,011	19,409	301	1416	73,819	72,333
1870-1899	7801	15,594	134	555	81,319	79,944
1900-1929	5823	11,455	21	100	87,008	85,676
1930-1969	6770	13,209	2	14	93,757	92,529
1970-1985	954	2401	0	0	94,709	93,498
Totals	97,209	96,739	2500	17,643		

(a) New Words, Obsolete Words, and Words in Use, by Period

Period	Mean	Median	Standard Deviation	Minimum	Maximum
1st to 12th centuries	8.10	0	97.23	0	2776
13th century	50.14	3	166.95	0	1064
14th century	165.61	13	473.26	3	4061
15th century	161.02	41.5	381.43	5	3091
16th century	358.10	308	269.11	28	1425
17th century	467.73	415.5	330.60	135	2638
18th century	217.35	174.5	138.93	76	971
19th century	605.59	593.5	189.34	260	1051
20th century	238.19	245.5	120.26	2	476
Totals	133.06	2	272.72	0	4061

(b) Average Annual Number of Words Added to the English Language, by Century

Lexical semantic change or diachronic linguistics deals with the change of language, specifically its meaning, over time. The aim for this research is to create a method to detect this change, preferably unsupervised for better generalized corpus. The corpus in use will be over 1,000,000 tweets over the past 10 years and the method of detection will be diachronic word embeddings into a recurrent neural network to predict the patterns within these changes.

## References

- [1] 1941 Crystal, David. *English as a global language*. Cambridge University Press, Cambridge, UK ; New York, 2nd ed.. edition, 2003.
- [2] Michael Reksulak, William F. Shughart, and Robert D. Tollison. Economics and english: Language growth in economic perspective. *Southern Economic Journal*, 71(2):232–259, 2004.