

# NLP Project 1, Part 1

---

Vahid Kharazi

November 13, 2015

## 1 CHURCH AND GALE

Yes, It's true. There is 15758 tokens and 2147 unique one with simple normalization(lower case all tokens). Without normalizing, there is 2421 unique tokens.

```
(venv)vahid@kharazi:~/dev/nlp/one$ python main.py q1  
|V| = 2147, N = 15758,  $N^{(1/2)} = 125.530873$ 
```