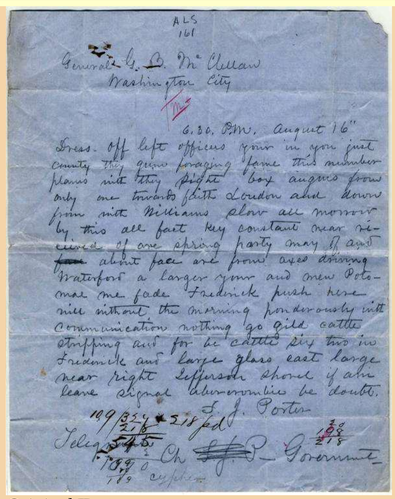
Cracking An Unsolved Civil War Cipher

Young Eun Suk & Becky Rovner

Professor Daniel Lopresti & Professor John Spletzer

Computer Science and Engineering, Lehigh University, Bethlehem, PA 18015



Original Text:
"Dress off left officers your in you just county they gum foraging fame this number plains with they sight box augurs from only one towards faith Loudon and down from with Williams slow all morrow by this all fact key constant near received of are spring party may of and about face are from axes driving Waterfont a larger your and men Potomac me fade Frederick push here will without the morning ponderously with communications nothing go gild cattle stripping and for be cattle six two in Frederick and large glass east large near right Jefferson shovel if am leave signal Abercrombie be doubt"

Civil War-era Military Document:

- Sender: F.J. Porter to Union General G. B. McClellan
- ~150 years old
- Possession: Lehigh University Library
- Status: Yet to be decrypted, original codebook unknown
- Previous work: Program by Professor Lopresti which output all possible decryptions of the letter.

Purpose of Research:

The goal of our project is to write a program that can grade texts based on their resemblance to proper English. This could be used to narrow down the possible decryptions of the civil war message output by Professor Lopresti's program.

Background

What is an encryption scheme?

The letter is broken up into a matrix and each word is a different element in the matrix. It is encrypted by transposing multiple columns or rows in the matrix in a specific manner. The order and manner in which they are shifted is the encryption scheme. The civil war message is encrypted with what is called the Stager cipher, which randomizes full words of a text using specific column/row routings. The diagrams to the right illustrate this method with the phrase: "Go to the river and move east for two miles. You will find another letter there." The encryption scheme is:

Third row, shift one element right.
Second column, shift one element down:

Go	To	The	River
And	Move	East	For
Two	Miles	You	Will
Find	Another	Letter	there

The encrypted message is then:

Go	Another	The	River
for	To	Move	East
Two	And	You	Will
Find	Miles	Letter	there

Technical Approach (Solution):

Develop a method to narrow down all the possible encryption schemes by using two primary programs that grade a body of text:

1. **Bigrams Program**:
   - Take a large text corpus as input
   - Tag all words as their appropriate parts of speech
   - Calculates and outputs the parts of speech bigrams statistics

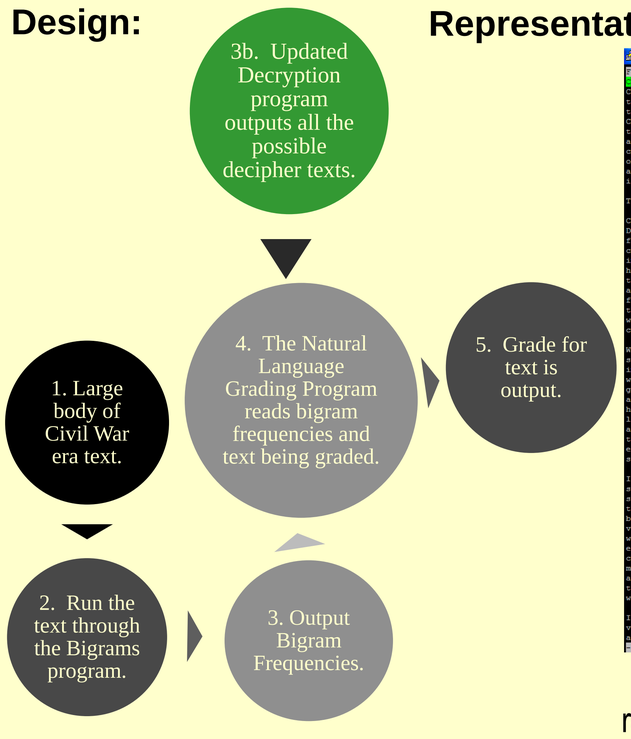
2. **Natural Language Grading Program**:
   - Takes the parts of speech bigram statistics as the input
   - Also takes a text to be tested as input
   - Calculates and outputs a grade based on the bigram statistics

(the grade is indicative of the text's resemblance to formal English)

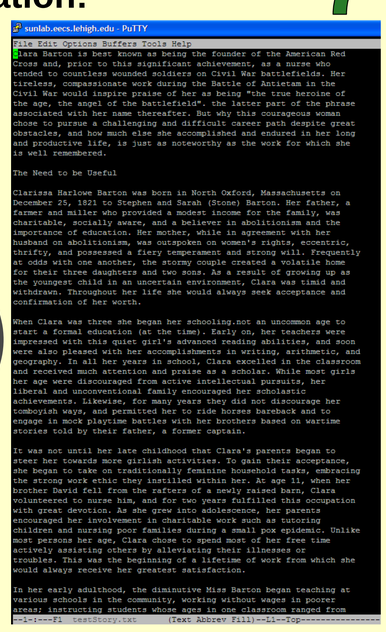
**Parts of speech tagger is courtesy of Stanford University.
Software: Stanford Log-linear Part-of-Speech Tagger www.nlp.stanford.edu/software/tagger.shtml

Implementation

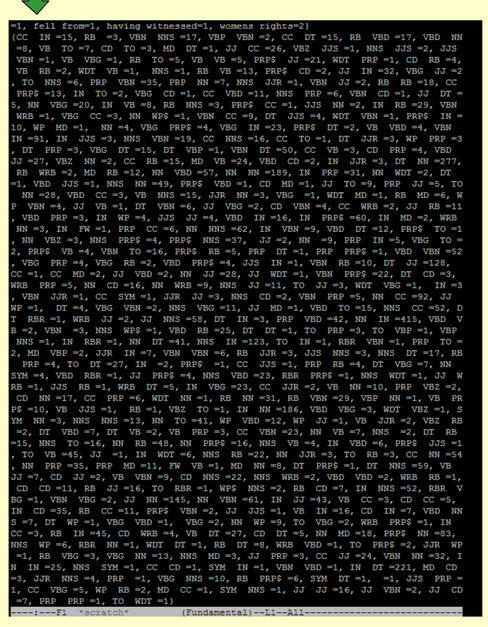
Software Design:



Visual Representation:



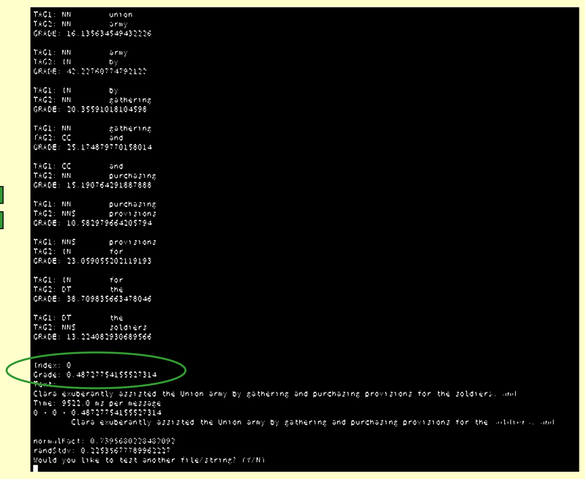
Bigrams Program that reads in large body of text.



Outputs a hash table with the bigram parts of speech frequencies.

Clara exuberantly assisted the Union army by gathering and purchasing provisions for the soldiers, and

Natural Language Grading Program reads in hash table and string of text being graded. (example string shown above)

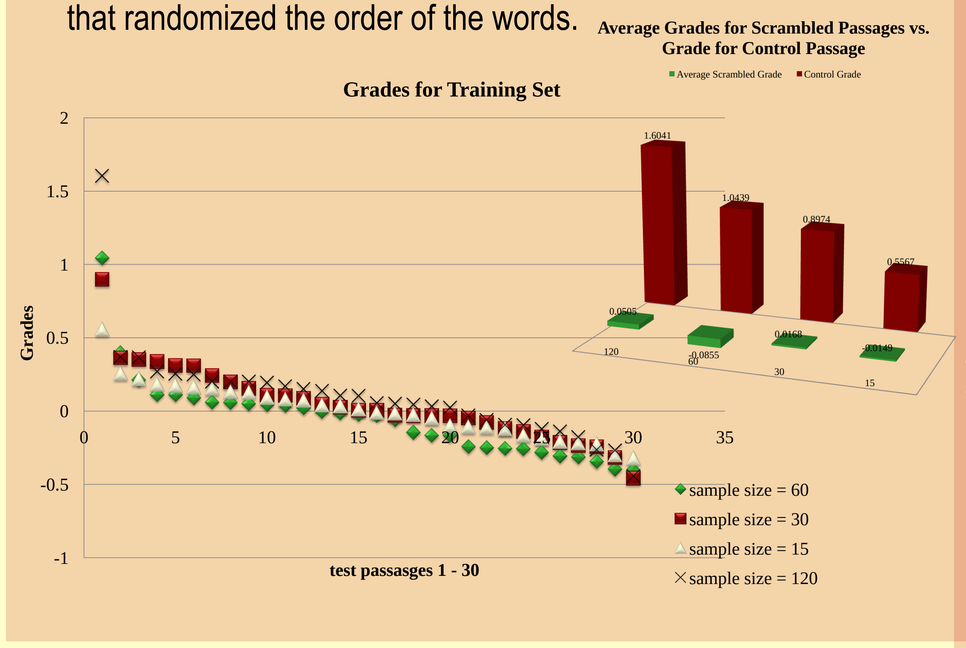


Natural Language Grading Program calculates a grade based on the frequencies stating how correctly the sentence is written. (After testing, the grade for a proper sentence is anything above 0.4.)

Testing

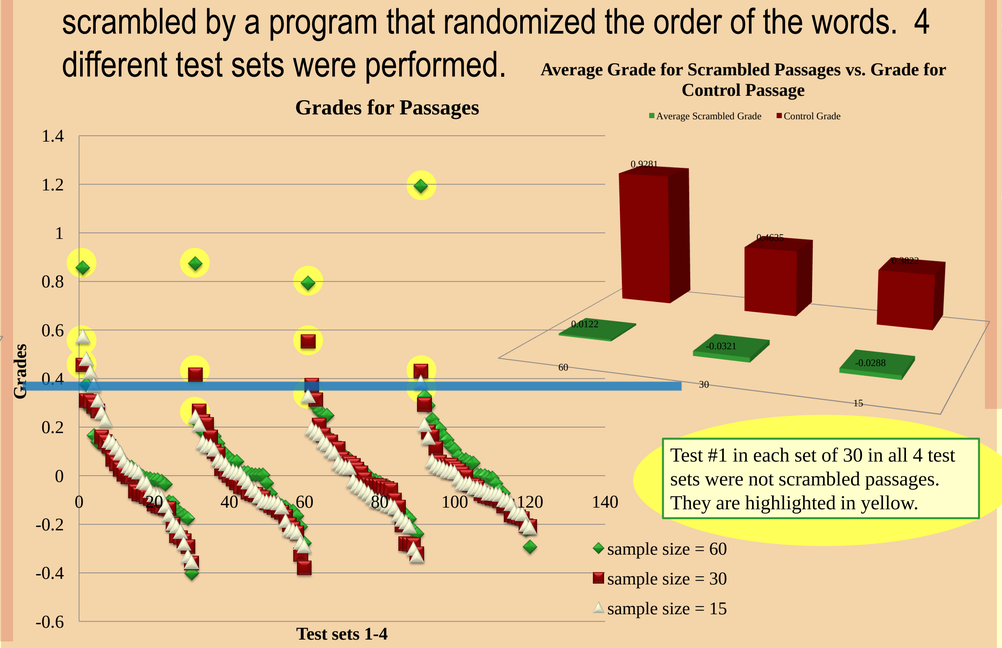
Training Set

15, 30, 60 and 120 word passages that could be found in the original bulk text were tested and given grades. 1 out of the 15, 30, 60 and 120 word samples were not scrambled and the rest of the passages were scrambled by a program that randomized the order of the words.



Testing Set

15, 30, and 60 word passages that could not be found in the original bulk text were tested and given grades. 1 out of the 15, 30, and 60 word samples were not scrambled and the rest of the passages were scrambled by a program that randomized the order of the words. 4 different test sets were performed.



Test #1 in each set of 30 in all 4 test sets were not scrambled passages. They are highlighted in yellow.

The next step would be to use the Natural Language Grading Program to narrow down the possible encryption schemes. To corroborate our solution a decryption of the letter from Queens University produced a score output by our NLG program of 0.4554, which is above our 0.4 threshold, demonstrating our program's effectiveness. Since the parameters of the cipher are still unknown for this proposed solution, there is further work to be done. However, there may be other Civil War military documents yet to be decrypted that could utilize our software to narrow down their possible encryption schemes.

Technical Requirements

Development Platform: Windows
Development Environment: Eclipse
Software: Stanford Log-linear Part-of-Speech Tagger
www.nlp.stanford.edu/software/tagger.shtml
Languages: Java and Tcl/Tk

Acknowledgements:

We would like to acknowledge the Lehigh University libraries and the "I remain" digital collection, especially librarians Ilhan Citak, Rob Weidman, Christine Roysdon, Julia Maserjian, and Phil Metzger. Without their help we wouldn't have had access to and a better understanding of the Civil War document. Also, we would like to express our gratitude to the Computer Science and Engineering Department for giving us this opportunity to share our work. Lastly, we would like to thank Professor Lopresti for his hard work, contribution and guidance, along with Professor Spletzer for his great advice and counseling. It was with their support and encouragement that made the completion of this project possible.

LEHIGH UNIVERSITY®