

# Syntactic Steganography

Hardik Rajpal

November 27, 2023

## 1 Introduction

Steganography refers to the concealment of information within a non-secret message or object. The technique has been applied in images and texts with goals of copyright protection and confidentiality. [Put reference papers here]. This project explores a form text steganography that exploits the syntactic redundancies of the English language. Given a sentence, it can be paraphrased to a different sentence, while retaining the (almost entire) meaning. Given a **plain document**, one can use the above facts to encode a **message** into it to produce an **output document**. To extract the message from the output document, the implementation requires the **plain document**.

## 2 Approaches

This section highlights the first few approaches that led up to the final implementation.

### 2.1 Syntax-Tree Manipulations

### 2.2 Hash-store Encoding

### 2.3 Predictable LLM Paraphrasing

## 3 Review

### 3.1 Steganography Dimensions

As highlighted in this paper, steganography techniques can be studied with three dimensions. Below I review the project with each dimension:

#### 3.1.1 Payload Capacity

It refers to the ratio of hidden information to cover information. The payload capacity is almost as good as methods of [Lexical steganography?], with the recurring example providing at most 2 bits per sentence. Additionally, the payload capacity is flexible and can be tuned based on the text provided.

#### 3.1.2 Robustness

It refers to the ability of the system to resist against changes in the cover object. [Discuss]

#### 3.1.3 Imperceptibility

the potential of the generated stego object to remain indistinguishable from other objects in the same category.

### **3.2 Limitations**

### **3.3 Takeaways**

## **4 References**

1. Linguistic Steganography Researchgate