A Survey of Different Methods of Linguistic Transformation for Steganography

Panpan Lin

A research paper submitted to the University of Dublin,

in partial fulfilment of the requirements for the degree of

Master of Science Interactive Digital Media

2014

Declaration

I declare that the work described in this research paper is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed: _____

Panpan Lin

28/02/2014

Permission to lend and/or copy

I agree that Trinity College Library may lend or copy
this research paper upon request.

Signed: _____

Panpan Lin

28/02/2014

ACKNOWLEDGEMENTS

Firstly I would like to thank my supervisor Prof. Carl Vogel, who first introduced me to the fascinating subjects of computational linguistics and steganography.

In writing this research paper I have benefited from the support of many other people, such as Alex Cairns, David Kelly, Vinayak Das Gupta, my classmates and housemates.

A special thank to my parents Baiheng Lin and Zhongcheng Tang.

Summary

This research is a survey of stegosystems based on different methods of linguistic transformation, with meaning preservation as the main concern. By discussing and comparing different systems, it is concluded that stegosystems based on shallow techniques and syntactic transformations perform better than systems based on deep parsing techniques and semantic transformation, but as NLP techniques develop, stegosystems based on semantic transformations will also have a bright future as they are capable of high payload capacity. Humans are the best way to evaluate a stegosystem in terms of its ability in preserving meaning. A common framework or benchmarks for evaluating linguistic steganography in terms of meaning preservation are needed.

Table of Contents

List of Tables and Illustrations

Introduction

This paper is a review of the different methods of linguistic transformation for steganography, focusing on meaning preservation. So far, there has been no systematic literature available on this topic, a gap the present paper attempts to fill. The language covered is mainly English, but some methods could also be useful for other languages.

Despite recent development in the field of linguistic steganography, preserving the meaning of the cover text remains difficult. Yet meaning preservation is crucial for the security of the stego system, and the key for future development in linguistic steganography (Bennett, 2004; Keith Winstein, 1998). Semantic attacks on linguistic steganography will become an issue, especially with human reader as potential detection mechanism for steganalysis, and improvements in semantic and rhetorical correctness in linguistic steganography becomes an important direction of future work (Bennett, 2004).

Research in linguistic steganography with a focus on meaning preservation is also likely to benefit other related technologies, such as digital text watermarking, machine translation, automatic text summarisation, and other areas in natural language processing.

Existing linguistic steganography methods and software tend to be text-based.[1] However, given the nature of languages, linguistic steganography does not have to be text-based and can be applied to languages in audio and video alike. The reason that existing linguistic steganography methods are mostly text-based can be easily explained by the state-of-the-art of NLP (Natural Language Processing) technologies. If voice recognition and generation system are further advanced, then many of the existing linguistic steganography methods could be applied to other media, notably for methods that generate their own cover datatype.

Chapter 1 provides some background information related to this research, including some basic concepts of steganography and linguistic steganography, and discusses existing reviews in linguistic steganography. Chapter 2 introduces and briefly discusses different existing systems of steganography based on linguistic transformations, with meaning-preservation as the focal criteria. Chapter 3 presents some thoughts on the systems discussed. Finally, a conclusion to the research is given, followed by discussion of limitation and possible future works.

[1] "Text" in this paper is used in its narrow sense, instead of the broader sense often adopted in literary theory.

Chapter 1

Background Information and Literature Review

1.1 Steganography

The term “steganography” can be traced back to Trithemus’s (1462-1516) work “Steganographia”, which derives from the Greek words στεγανός and γραφή, meaning "covered writing" (Petitcolas, Anderson, & Kuhn, 1999). Kahn (Kahn, 1996) presented a comprehensive account of historical examples of steganography at the first international conference on information hiding. Most scholars cited Herodotus’ accounts as examples (Bennett, 2004), most particularly that relating to how Histaeus sent a secret message by shaving and tattooing the message on his slave, then waited for the slave’s hair to grow back before sending him off to his destination. Another commonly cited account from Herodotus is how Demeratus sent a message from the Persian court back to Greece. He concealed the message by carving it on a wooden board, then covering it with wax, thus giving it the appearance of how a blank writing board looked like at that time. Another example mentioned by Kahn and frequently cited by scholars is invisible ink, which has been used since ancient times and was still notably during World War II.

Being different from cryptography, which tries to encipher the secret message carried, steganography tries to hide the very fact that there is secret communication going on. A classic illustration of steganography is “the prisoners’ problem” proposed by Simmons (Simmons, 1984): Two prisoners, Alice and Bob, are put in separate cells. They would like to coordinate an escape plan. All their communications will have to pass through and be examined by the warden. If the communications cause the warden to suspect that there is a conspiracy, communication can be cut off and the escape becomes impossible. Under this circumstance, a crypted text that does not make sense is very likely to arouse the warden’s suspicion and it is therefore better to use steganography to produce some seemingly innocuous texts that in fact hide secret messages. Any text allowed to be passed between Alice and Bob can serve as a cover text, and if it is imbedded with hidden information, it is called a stego text. The datatype used for hiding message is not restricted to text, but can be image, audio etc. The algorithms involved in embedding the hidden information into the cover datatype and in retrieving the hidden message from the stego datatype form a stegosystem (Pfitzmann, 1996). Nevertheless, hidden messages are usually encrypted first before being hidden in the cover datatype to achieve better security (Bennett, 2004).

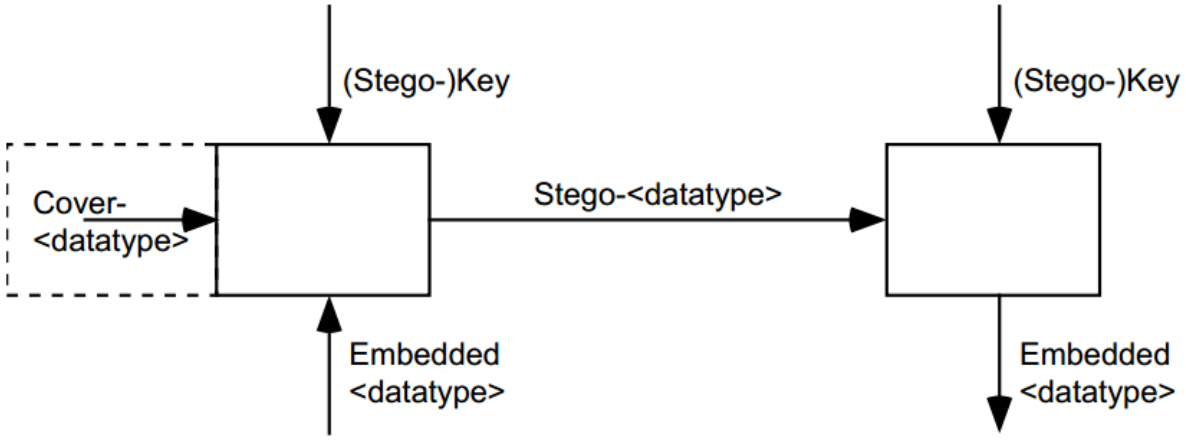

Fig 1 (Pfitzmann, 1996)

Steganography is also closely related to digital watermarking, but unlike watermarking, steganography's main aim is to hide the existence of secret messages from outsiders, while digital watermarking deals with authentication, such as copyright, and aims at marking a document with information that is robust against modifications (Shih, 2007). Moreover, in digital watermarking, the information and the document being marked are closely related, while in steganography, the cover text and the hidden message are usually unrelated.

Johnson and Ktzenbeisser (Johnson & Katzenbeisser, 2000) group steganographic techniques into six categories by how the algorithm encodes information in the cover objects: substitution systems, transform domain techniques, spread spectrum techniques, statistical methods, distortion techniques, and cover generation methods.

Steganalysis is the art of discovering the existence of hidden information in communication, retrieving the hidden information if possible, and distorting or even removing the embedded message. Wayner (Wayner, 2009) classifies stego attacks into three types: visual and aural attacks, structural attacks, and statistical attacks. Visual and aural attacks describe the human factor in stego attacks, which is most relevant to this paper. In steganography there exists a trade-off between security and payload (Chang, 2013). A stegosystem capable of embedding a lot of hidden information with a relatively small cover datatype has a high payload, and is efficient. However, when a large amount of information is to be embedded into a cover datatype, it can make the cover datatype noisy or statistically aberrant, making it very suspicious to a human being or computer and, thus provoking an attack.

1.2 Different Types of Steganography

With the development of computer technologies, digital media such as images, video, and audio are used as cover media for steganography (Fridrich, 2009). Digital steganography systems usually exploit the redundancy of the cover media, and rely on the limitations of the human sensory systems (Chang, 2013). With some media, it is easier to achieve a higher security level and lower payload capacity, making them more popular than others media. Images, for example, are quite popular as cover media due to their abundance on the Internet and the large payload they can achieve. Image steganography usually takes advantage of the naturally occurring noise in an image. A common way of embedding secret messages in images is to change the least significant bit (LSB) in the colour representation, as human beings are not likely to be able to detect this. Similarly, audio steganography seeks to hide secret messages by exploiting the frequencies that humans cannot hear (Johnson & Katzenbeisser, 2000). However, these methods lack robustness and a simple compression can destroy the hidden information.

Text steganography uses text as the medium to hide secret information, for example by employing text reformatting, lexical substitution, syntactic transformation, and cover text generation (Bennett, 2004). Bennett groups these techniques into three categories: format-based methods, random and statistical generation, and linguistic methods. However, it should be pointed out that although linguistic steganography is a way of implementing text steganography, linguistic steganography does not have to be text-based. Speech steganography can also be linguistic steganography. Both format-based methods and random and statistical generation methods are non-linguistic. An example of the format-based method is: imbedding secret information by inserting spaces or by changing fonts. Random generation steganography can be easily detected by either a human being or a computer capable of carrying out a statistical analysis of the stego text. Statistical generation, such as Wayner's (1992) method of using the Huffman compression algorithm in reverse to generate a cover text based on an existing natural text, can produce a stego text that has a much greater resemblance to a natural text, but it is still easily detectable, especially for a human being, as the stego text generated is nonsensical. Both random generation and statistical generation have serious drawbacks and are not really practical for steganography (Bergmair, 2004; Chang, 2013). Linguistics methods, which will be discussed in more details in section 2.3, are linguistically driven in generating or modifying cover text, resulting in cover texts with higher readability and leading to a more secure system.

1.3 Linguistic Steganography

Chapman, Davida, and Rennhard (Chapman, Davida, & Rennhard, 2001) defined linguistic steganography as "the art of using written natural language to conceal secret messages". This definition has two problems. First, methods such as reformatting the text or changing fonts are using natural language to conceal secret messages, but do not involve linguistics knowledge. Second, linguistic steganography does not necessarily rely on text. It can be based on other media as well. This paper adopts the definition of Bennett (Bennett, 2004), Bergmair (Bergmair, 2007), and Chang (2013) uses in their research, which is narrowed down to steganography that specifically operates on linguistic properties, and excludes techniques that rely on superficial properties of text or on specific file formats, such as techniques that directly operate on typeset or written text as a graphic, or speech as a waveform.

According to the origin of the cover text used, linguistic steganography systems can be classified into two categories: they either train the computer to generate a meaningful, natural text to be used as the cover text in which to hide messages (generation-based), or take an existing text as cover text, and to apply linguistic transformations to the text to produce a stego text that is meaningful and natural (transformation-based). This paper focuses on transformation-based methods.

Some examples of linguistically driven generation methods are the context-free grammars (CFGs) system developed by Wayner (Wayner, 2009), the NICETEXT system developed by Chapman and Davida (Chapman & Davida, 1997), the Listega (Desoky, 2009a), Notestega (Desoky, 2009b), and Edustega (Desoky, 2011) developed by Desoky.

To achieve a high security level, beyond not showing significant statistical anomalies, it is very important that the linguistic steganography methods produce a stego text that is meaningful and natural for a human reader, both from a grammatical point of view, and from a lexical and contextual point of view. Cover texts generated based on Wayner's CFGs are grammatically correct, but senseless and lack semantic consistency, making it easy to be spotted by humans. Chapman and Davida's method takes semantics into consideration, along with syntax, and uses style sources and large code dictionaries that categorise words semantically. They later improve the system by using synonyms and extensible contextual templates (Chapman et al., 2001). As a result, the cover text generated by their system has a higher readability then the CFG method, but it is still far from natural for a human being. For

NICETEXT to generate a good quality text with little repetition that also makes sense, it is crucial, apart from having a large number of templates, that it also uses a well written synonym dictionary. Even if statistical methods and n-gram are introduced to improve synonym substitution, the large number of templates required to generate contextually acceptable texts still makes this method quite unscalable. Given the current state of NLP, linguistic stego system relying on generation method usually restricts itself to generating highly specific texts with a limited structure and a specific vocabulary, such as spams (Spammimic.com), lists, and notes. Stego systems based on linguistic transformation, which will be discussed in details in the rest of this paper, are thus more practical for the time being as they use existent texts as cover texts.

Due to the naturally low level of redundancy in natural language, linguistic steganography has a lower payload compared with many other forms of steganography based on digital media, such as image and audio steganography based on non-linguistic methods (Cox & Miller, 1997). However, linguistic steganography can exist outside the digital realm, and is more robust against digital attack. Text steganography based on non-linguistic methods can easily be destroyed by transcription, text reformatting, or OCR scanning (Atallah et al., 2001), but none of these methods can destroy a stego text achieved by linguistic steganography.

Chang (2013) has summarised the encoding methods adopted by linguistic steganography: block code method, mixed-radix number method, Huffman code method, and hash function method. The block code method and the mixed-radix method give the same probability to each representation, which can cause a security issue since some transformations may be preferred by native speakers over other transformations. The Huffman code method is a variable-length code method, allowing more frequently used expressions or structures to be chosen over those used less frequently. Unlike the previous three methods, the hash function method is independent of the cover text and is thus favoured by machine-translation based stegosystems. In addition to these, Chang adopts the vertex colouring coding method in her own lexical substitution-based stegosystems (Chang, 2013), which encode message by turning synonym graph into a coloured *k*-chromatic graph so that each word is encoded by the code that represents the colour. The chromatic number of each graph is decided by the size of the synonym set, which is set to 4 in Chang's system, as her system is based on WordNet (Fellbaum, 1999), whose synonym sets usually do not contain more than 8 synonyms in each set (Chang, 2013). The current paper focuses on linguistic transformation for steganography rather than their encoding methods.

1.4 Literature Review

1.4.1 Bennett (2004)

Bennett's survey on linguistic steganography focuses on cover generation methods (Bennett, 2004), while this paper focuses on linguistic transformation methods. However, Bennett has highlighted some of the problems inherent in text steganography and problems with different linguistic approaches. Although systems using the Huffman compression algorithm in reverse can withstand statistical attacks and system with syntactical templates produce cover texts with correct grammars, all the systems Bennett has analysed suffer due to their lack of senses and readability. Bennett points out that natural language cover texts must not only pass statistical analysis, but also the minds of human readers. Therefore the meaning and semantic cohesiveness of the cover text will become more and more important and should be a future research direction.

1.4.2 Bergmair (2004)

Bergmair (Bergmair, 2004) examines both theoretical approaches and practical implementations of linguistic stegosystems. Four linguistic steganography systems are analysed: Winstein's Tyrannosaurus Lex (K Winstein, 1999), Chapman and Davida's NICETEXT (Chapman & Davida, 1997; Chapman et al., 2001), Wayner's mimicry systems (Wayner, 1992) including Spammimic, and the syntactically-based system developed by Atallah et al. (Atallah et al., 2001; Atallah et al., 2003). Among them, the systems developed by Winstein and Atallah et al. are most relevant to this research. Winstein's system is based on lexical substitution and Atallah et. al.'s system is based on syntactic transformation, though they would later go on to develop another system based on semantic transformation. Both Winstein's and Atallah's stegosystems strive to preserve meaning.

1.4.3 Topkara et. al. (2005)

Topkara et. al. (2005) provides an overview of natural language watermarking techniques and tools, compares and contrasts natural language watermarking with image watermarking. Though the topic is on linguistic watermarking instead of steganography, many of the issues covered are overlapped.

In their survey, characteristics of three categories of linguistic transformation for text watermarking are identified and discussed: synonym substitution, which is most widely used,

syntactic transformation, including changing syntactic structure and verb-based transformation such as locative alternation, and semantic transformation, especially using noun phrase coreferences for coreferent pruning, grafting, and substitution. They point out that the biggest problem with synonym substitution is computers' poor performance in word-sense-disambiguation, while for semantic transformation, the problem lies in the fact that coreference resolution is one of the hardest tasks in natural language processing. They believe the most promising approach to be embedding information in the syntactic structure of sentences, which falls into the second category, due to the quality of syntactic analysis tools available, but similar to Bennett (2004), they also point out the importance of coherence of semantics and rhetorical structure and mention them as a direction for future research.

As to specific technologies in steganography, they look at the probabilistic context-free grammar approach, the synonymy substitution approach, and the text-generation approach using hybrid techniques represented by NICETEXT. They draw to the same conclusion as other researchers about the impracticability of using probabilistic context-free grammar as the texts produced is ungrammatical, nonsensical, and so only useful for situations where only a computer is acting as the arbitrator.

They believe that techniques modifying the underlying structure of a sentence will be more robust than those that modify only the surface representation of the sentences, such as Atallah's methods (Atallah et al., 2001; Atallah et al., 2003) which embed message by changing the syntactic and semantic trees of a sentence.

They also provide a summary of relevant resources, tools, and techniques. Resources include large corpus for statistical analysis and machine learning systems, such as the Penn Treebank, and electronic dictionaries designed as large databases of lexical relations between words, such as WordNet. Relevant tools listed are parsers, generators, and semantic analysers. Techniques involved are parsing, including POS tagging, natural language generation and text paraphrasing.

They identify natural language watermarking as a specific kind of machine translation, which translate between the same language. They identify grammaticality, meaning preservation, and text style as key factors to consider when evaluating if a linguistic watermarking system is successful or not. They conclude that evaluation system for machine translation, represented by BLEU, can be adapted for evaluation for linguistic watermarking system. However, since computers are still very ineffective in word-sense-disambiguation and

coreference resolution, which they have mentioned before when analysing the three categories of linguistic transformation, machine translation evaluation system is not likely to produce satisfactory in evaluation, especially for evaluating meaning preservation, which happens to be one of the key factors in such evaluation.

As the survey aims at giving a general introduction to the subject, it does not cover many stegosystems, and discusses general techniques without going into specific stegosystems, which will be covered by this research paper.

Chapter 2

Stegosystems Based on Linguistic Transformation

Linguistic transformations adopted by existing stegosystems can be classified into three categories: lexical or phrase substitutions, syntactic transformations, and semantic transformations, with the first two categories being more common (Chang, 2013). However, when trying to categorise the following systems into these three groups, I have come across some difficulty. There are some methods which take into consideration elements from different categories. Thus in the end, a historical approach is adopted instead, resulting in the systems arranged according to the year in which they are developed.

As Chang has observed, substitution-based methods have the advantages of tending to be grammatically correct, but present challenges in word-category disambiguation, word-sense disambiguation, and semantic and pragmatic differences among synonyms; syntactic transformations, such as passivization/activization, topicalization, and clefting, involve modifying the syntactic parse tree and reconverting the deep structure into another surface structure. Since the current research focuses on meaning preservation, these machine-translation based methods will not be considered.

2.1 Wisenstein's *Tyrannosaurus Lex* (1999)

Tyrannosaurus Lex (K Winstein, 1999) is a lexical steganography system based on synonym substitution and with a focus on meaning preservation. The system makes use of existing synonym sets in WordNet. In the original cover text, words that appear in one of these synonym sets are used to encode bits. The encoding method is block coding and different words in the set are assigned different values depending on their position. The decoding process does not need the original cover text, but only the stego text and the synonym sets.

However, as Winstein (1998) has observed, the system does suffer from several drawbacks. Beyond the lack of adaptability and inefficiency in encoding bits, there exists a security issue due to the lack of a mechanism for word-sense-disambiguation, resulting in the possibility of substitutions that are very likely to cause suspicion from a human reader. Given the large number of polysemous, this can be a serious problem. In addition to this, the use of block coding does not allow bias towards words that are more likely to be a more suitable substitution for the original word, thus lowering the security level of the stegosystem.

To solve the problem in word-sense-disambiguation, Winstein defined "interchangeability sets", which are sets of words that appear only in the same synonym sets. A database of approximately 20,000 words distributed in different interchangeability sets is created. However, as Chang (2013) has pointed out, this approach will result in a considerable reduction in the payload capacity as many synonyms will have to be discarded, while the security issue still exists due to the fact that many synonyms are only applicable in certain contexts.

2.2 Atallah, Raskin et. al. (2001, 2003)

The systems developed by Atallah, Raskin et. al. (Atallah et al., 2001; Atallah et al., 2003) are not designed for linguistic steganography, but natural language watermarking, which focuses more on the robustness of the system. However, since Atallah's system embeds secret message by applying syntactic and semantic transformation to an existing text while preserving its overall meaning, it is highly relevant to steganography based on linguistic transformation.

As Bergmair (Bergmair, 2004) points out, one of the important aspects of Atallah's approach is that it takes an explicit approach to meaning by representing semantics in Artificial Natural Language (McCarthy, 1990). The essence of their approach is very similar to that of machine translation.

The theoretical foundation of Atallah's systems is ontological semantics (Nirenburg & Raskin, 2001, 2004), which is "a theory of meaning in natural language and an approach to natural language processing (NLP) that uses a constructed world model, or ontology, as the central resource for extracting and representing the meaning of natural language texts and for reasoning about knowledge derived from texts" (Nirenburg & Raskin, 2004). This theory is more a vision of an ideal system rather than one that could actual be implemented, but has already been adopted by many researchers in machine translation. Atallah et. al. have adapted this theory to their own system. The ontological semantic system contains a set of static knowledge sources, knowledge-representation languages, and a set of processing modules (Nirenburg & Raskin, 2004). In terms of the static resources, Atallah et. al.'s system is concerned with the ontology, the lexicon, and the text-meaning representation (TMR). As to the dynamic resources, they include a syntactic parser, an analyser, and a generator. The ontology is a collection of all the concepts in a given world, including entities, events, and their relationships, arranged in a tangled hierarchical order. The lexicon contains lexical

entries and in the case of machine translation, the lexical entries will be taken from different languages. The TMR of a sentence is a list of prepositions representing the meaning of the sentence and is constructed by mapping its lexical items in relation to their ontological concepts. The parser and the analyser are used in turning the original text into its TMR, while the generator and a reversed parser are used in turning the TMR to the watermarked text, or in the case of steganography, a stego text.

2.2.1 Syntactic Transformation-Based System

However, the system first developed by Atallah et. al. (2001) is based on syntactic tree structure rather than the TMR trees. They used the Link parser (http://www.link.cs.cmu.edu/link/) developed at Carnegie Mellon University, then apply syntactic transformations that have been proven to be effective in preserving meanings, namely adjunct movement, clefting, and passive formation to chosen sentences. The following table (fig 2) shows some sentences taken from a text that has been processed by the system. The text is provided by Atallah et. al. (http://projects.cerias.purdue.edu/wmnlt/demo/index.php):

Original Sentence	**Transformed Sentence**	**Transformation**
The functions of these instruments are discussed in the Appendix.	In the Appendix the functions of these instruments are discussed.	Adjunct movement
EOD teams have equipment for detection of gaseous radioactivity.	It is EOD teams that have equipment for detection of gaseous radioactivity.	Clefting
The shore installation support will be contingent upon the location of the accident and the ability to continue with the primary mission of the shore installation.	Specifically the shore installation support will be contingent upon the location of the accident and the ability to continue with the primary mission of the shore installation.	Adding adverb
Ships in the vicinity may provide some additional assistance.	Some additional assistance may be provided by the ships in the vicinity.	Passivisation

Prepare to initiate DECON station procedures.	Initiation of DECON station procedures shall be prepared.	Passivisation (with nominalisation)
The CO is responsible for informing the ship's crew regarding PA releases.	It is the CO who is responsible for informing the ship's crew regarding PA releases.	Clefting
The most important thing is to perform operations to minimize hazards to the ship's personnel and damage to critical equipment.	To perform operations to minimize hazards to the ship's personnel and damage to critical equipment is the most important thing.	Clause switching

Fig 2

Judging from the sample, Atallah's system generally performs well in meaning-preservation and could be useful for steganography purposes. However, if the system is to be transformed into a stego system, there are still some issues that might cause suspicion either from a human reader or a machine.

Firstly, some syntactic transformations, such as clefting, are not used very often in natural text. If a higher level of security is to be achieved, a statistical parameter can be introduced into the system to control the frequency of cleft sentences in the text so that this percentage falls within the normal range of cleft sentences in natural texts of the same genre. However, the payload is then likely to be reduced. A possible way to compensate for the reduction in payload capacity is to introduce a reverse-cleft method, which can be applied to existing cleft sentences in the cover text. This method has the double advantage of being able to embed secret bits on its own and reducing the frequency of cleft sentences in the text to normal, thus allowing clefting to be performed on other sentences in the text without drastically increasing the frequency of cleft sentences in the cover text.

Secondly, there are issues with contextual cohesion. For example, in the fifth example given above, the original sentence "prepare to initiate DECON station procedures" is situated in the following context: "Near shore releases should be done as a last resort action. Attempt to save the lives of personnel. Prepare to initiate DECON station procedures. Initiate initial Operations Report (OPREP)-3." As can be seen, the sentence to be transformed is situated within a series of imperative sentences which, together, form an instruction to follow. The

transformation suggested: “Initiation of DECON station procedures shall be prepared” does not fit in well within this context. A tool is needed to check the contextual cohesion of the transformation proposed. In this specific case, the issue can be detected by comparing the parse trees of the sentence to be transformed with its surrounding sentences.

This system can also cause problems in stylistic cohesion. Atallah et. al. have also mentioned that their approach works best with meaning-oriented texts like reports and manuals rather than stylistically rich texts, such as literary works.

This system has a low capacity with a bandwidth of 0.5 bit per sentence, which is drastically increased to 8 bits per sentence in a later system developed by Atallah, Raskin, Topkara et. al. (Atallah et al., 2003).

2.2.2 Semantic transformation-Based System

The new system is based on semantic transformation rather than syntactic transformation and functions by manipulating the TMR tree, which is much larger and richer then the syntax tree and provides multiple choices for each transformation suggested.

The system first creates TMR trees. The event proposition of the speech act of every sentence is chosen as the root, with filled slots of a concept suspended from it as branches (Atallah et al., 2003). Manipulation of the TMR tree is achieved by grafting, pruning, and substitution. The following table (fig 3) presents some examples given by Atallah et al (2003).

Transformation (resource used)	**Original Sentence**	**Transformed Sentence**
Pruning (co-reference)	The Pentagon ordered two new spy planes, including the unmanned “Global Hawk”, to the region to start flying *over Afghanistan*.	The Pentagon ordered two new spy planes, including the unmanned “Global Hawk”, to the region to start flying.
Grafting (co-reference)	The Pentagon ordered two new spy planes, including the unmanned “Global Hawk”, to the region to start flying over Afghanistan.	The Pentagon ordered two new spy planes, including the unmanned “Global Hawk”, to the region to start flying over Afghanistan, *which they are*

		attacking.
Adding/ Substitution (fact database)	The United States are attacking Afghanistan.	The United States are attacking Afghanistan, *which is ruled by Mullah Mo-hammed Omar*.

Fig 3 (Atallah et al 2003).

Sometimes there exist more than one way of transforming a single sentence. Judging from the examples given, Atallah's system has the ability to produce natural and coherent stego texts. By manipulating the semantic tree of a sentence, this system does alter meaning at sentence level, but on a deeper textual level, it is "meaning-preserving". Some precautions are taken to try to preserve coherence on a contextual level while carrying out the semantic transformation at sentence level: pruning is restricted to repeated information as witnessed by co-reference; in grafting, new propositions observed in co-reference are grafted on to the concept of the main tree; in adding/substitution, additional information is taken from the fact database to be added to the sentence to be transformed.

The two systems discussed are different from other systems in that they both function on a deeper level than many other systems, either by turning a surface structure into an ANL, from which they generate another surface structure, or by turning sentence meaning into TMR, which they manipulate according to co-reference and a fact database and from which they carry out transformation on the sentence. Moreover, Atallah. et. al. identify information in the co-reference and fact database as potential "noise" in natural language material. Since the size of the co-reference and the fact database are much larger than any syntactic structure, especially in the case of a fact database, which is "open" rather than "closed", this idea presents an opportunity for greatly increasing payload capacity in the future.

2.3 Bolshakov (2004)

Like Winstein's *Tyrannosaurus Lex*, Bolshakov's system is also a synonym substitution-based system that aims to preserve meaning and linguistic correctness (Bolshakov, 2005; Bolshakov & Gelbukh, 2004). However, in opposition to Winstein's system, which tries to improve the quality of the substitution by limiting the synonym sets to "interchanegeability sets", Bolshakov's system applies transitive closure to overlapping synonym sets, which will merge all the overlapping synonym sets into one large set. Although this approach avoids the decoding ambiguity, the synonym sets can also contain words that are not really synonyms of the word in the original cover text, thus reducing the level of security (Chang, 2013). To

solve this problem, groups of synonyms in Bolshakov's system are classified into two categories: absolute synonyms and relative synonyms, with different algorithms developed for each. Absolute synonyms are used independent of context, while relative synonyms need to pass a collocation-based test beforehand.

To implement this, and beyond the use of a specifically written synonym dictionary, a database of collocations is needed to verify the relative synonyms. The synonym dictionary is based on WordNet and is designed to have one dominant word within each synonym group, with the dominant word's absolute synonyms in the group labelled. Multiword and grammeme can also be part of the group. Since words that are not really synonyms of the original word in the cover text will still have the potential to be adopted as synonyms if they pass the collocation-based test, the security level of this system depends heavily on the quality and size of the collocation database. The collocation database used in this system is built from statistics collected from the Google search engine by sending both quoted and non-quoted queries of words to the engine. The mean value of both the quoted and non-quoted measurements is taken, and if this value is lower than a certain threshold, the synonym involved is not used for substitution (Bolshakov & Gelbukh, 2004).

As this method involves an attempt to quantify how natural different synonyms are within the same synonym set, which is very different from Winstein's system, one possible way of improving this system is to use variable-length code methods such as the Huffman coding method instead of the block coding method. In this way, collocations with a higher frequency can be favoured, increasing the chance of attaining better-suited substitution.

2.4 Topkara et. al. (2006)

2.4.1 Equimark: A Synonym Substitution-Based System

This synonym substitution-based method proposed by Topkara et. al. (2006) was originally designed for text watermarking instead of steganography, thus focusing more on the system's robustness. By replacing the targeted word with the word with the maximum ambiguity among all the synonyms of the targeted word, the system makes the transformation very difficult to reverse. Another feature of Topkara's method is that it continues to modify the document until the very end, so as to make it even more difficult to reverse. Such an approach is obviously taken from a watermarking point of view, but may also be used to

increase the security level of a linguistic steganography system, provided that it will not drastically change the statistical features of the text.

The synonym sets used for the system come from WordNet, and synonym sets related to different senses of the targeted word are used. The library WordNet::Similarity is used to find relevant semantic relations between synonym sets such as "is-a-kind-of" and "is-a-part-of" relations.

By introducing a more ambiguous word, a part of the more specific meaning of the original word is likely to be lost, thus such a transformation does not preserve the meaning completely. Yet, considering the inclusive nature of a hypernym to its hyponyms, it is possible that such a transformation may still preserve the meaning of the original word to some extent. The authors have proposed to use generalisation substitutions for meaning-preserving transformations, which means replacing the specific with the general, such as replacing *roses* with *flowers*. Since in a WordNet hierarchy, there can be more than one parent, the authors propose that it is possible that, given a certain context, they cannot move up one link due to some other words presented, but are able to move up another link. The example they give is to generalise *kangaroo* in a text that also mentions camels to *marsupial* instead of *herbivore*, since the latter can be understood as denoting the camel in this circumstance.

However, from the example above, the complexity of this method can already be seen. The problem with such a system is that, due to the poor performance of computers in dealing with meaning on different levels, it can be difficult to achieve a high level of accuracy, especially in a text that contains many hyponyms of the same hypernym. Besides, replacing *kangaroo* with *marsupial* in a children's story is not likely to get very far in terms of avoiding suspicion, which means that other factors, such as register and genre may also need to be considered. In this specific case, checking the word-frequency of the substitute may also be helpful. In addition, to achieve high accuracy, depending on the nature of the text used, it can become necessary to produce customised handcrafted synonym sets for each text, which has also been suggested by the authors themselves. Other solutions proposed by the authors include to use a more powerful ontology and to use a more domain-specific difference function that is also corpus-based, instead of relying solely on the WordNet-based difference function. All in all, even if such a system is made, it does not seem to be scalable or portable.

2.4.2 Syntactic Transformation-Based System

Topkara et. al. (2006) proposed a style- and meaning-preserving sentence-level watermarking scheme that embeds secret messages by syntactically altering the sentence structure. Similarly to many other methods based on syntactic transformation, such as Meral's scheme, Topkara's method also functions by converting a sentence into its structural representation by using a parser, which will produce a parse tree and derivation tree (dependency tree), after applying the relevant transformation on the parse tree and converting it into a deep syntactic structure, which will then be converted back into its surface sentence form. For this system, they use the Reuters corpus for testing, adopt the XTAG parser for parsing, and use RealPro for the final generation process.

The syntactic transformations involved are: passivisation, activisation, adjunct movement, and topicalisation.

For the evaluation, instead of using human beings to check for the naturalness of the sentences, they adopt the MT Evaluation Toolkit of NIST to evaluate the quality of the transformed sentences, which checks how close each transformed sentence is to the original sentence and outputs scores for BLEU metric. From a meaning preservation perspective, such an evaluation method may not be the best for providing feedback on the quality of the sentence. They also admit that using BLEU for sentence by sentence distance evaluation "is neither sufficient nor accurate" for this task (M. Topkara et al., 2006) : 44).

2.5 Vybornova & Macq's Presupposition-Based System (2007)

Simillar to Atallah et al.'s (2003), Vybornovoa and Macq's (2007) system is also based on semantic transformation, and claim to preserve meaning as well as grammaticality. In addition to this, both systems are just proof-of-concept prototypes.

Vybornovoa and Macq's system is based on presupposition, information conveyed in implicit form and often taken for granted by the readers. They classify seven presupposition triggers: definite noun phrases (NPs) and proper names; possessives; interrogative words; factive predicative constructions; implicative verbs; aspectual verbs and modifiers; it-clefts and wh-clefts. They give illustration and examples of definite NPs and Proper Names, Factives, Implicative verbs, aspectual verbs and modifiers, It-clets and wh-clefts.

For definite NPs and proper names, the presuppositions can be removed by removing the presuppositions trigger and paraphrasing the sentence.

For factives, which can be identified by predicates presupposing the truth of the information, such as *know that* and *realize that*, they can be replaced by a sentence with a non-factive, such as *believe* and *think*. Implicative verbs are verbs that are semantically rich and thus contain a large amount of presuppositions. For example, a sentence like "she managed to finish her homework" would have two suppositions: *she tried to finish her homework*, and *she did it*. It is suggested that the sentence can be changed by preserving only the result of the sentence, and in the example cited above, that means the second presupposition.

For aspectual verbs and modifiers, which can be identified by verbs and adverbs denoting the beginning, end or continuation of an action, it is suggested that the aspectual modifier can be removed while introducing the information presupposed by it. The same applies to it-clefts and wh-clefts. They presume the truth of the information contained in their subordinate clause, and can be transformed by removing the corresponding contructions while keeping the information involved.

As we can see from the above description, there are several problems with this system. Firstly, NLP techniques are not effective and accurate enough for this kind of deep semantic analysis, thus it is doubtful if such a system can become practical on a large scale in the near future. Secondly, it requires hand-crafted dictionaries or sets and thus not very scalable. Thirdly, it does alter the meaning of the sentence to some extent, for example in the case of implicative verbs, if only one of the presupposition is preserved, some meaning can be lost, even though one could still argue that in this case, a human reader will still find the meaning natural and nothing lacking and thus not affecting the security level of the system despite the fact that the meaning is slightly altered.

2.6 Murphy & Vogel (2007)

Murphy and Vogel (2007a, 2007b) propose a reliable meaning-preserving stegosystem based on relativiser swapping and complementiser swapping, as well as a system for text watermarking based on three statistically-constrained methods: lexical substitution, adjective conjunction swappping, and relativiser switching. Their approaches rely on parsing and part-of-speech (POS) tagging, which are mature techniques with high accuracy, thus leading to more accurate results and higher security levels.

The relativiser swap in Murphy and Vogel's stegosystem (2007b) embeds secret messages by dropping or changing the relative pronouns between *that*, and *who* or *which* in restrictive relative clauses. One difficulty lies in distinguishing human from non-human nouns, which is essential for deciding whether to use *who* or *which* in the restrictive relative clause. To solve this problem, an algorithm using WordNet and COMLEX is developed to measure how human-like the noun phrase in question is. The threshold is set at 0.5, which can be raised if a higher level of security is needed. The complementiser swap embeds message by dropping or inserting the complementiser *that* between the verb and the complement clause in verb complements.

As is mentioned in the previous section on Atallah et. al.'s method, shallow syntax-based methods may have a smaller payload capacity compared with the deep semantic methods, but as parsing technique is well developed compared with deep structure techniques, with modern syntactic parsers achieving around 90% accuracy per node, Murphy and Vogel's approaches are inherently advantageous in terms of meaning preservation compared with approaches depending on less mature deep structure techniques, thus achieving a higher security level. The reliability rate reported reaches 96% (Murphy & Vogel, 2007b). Not only do these two approaches discussed above preserve the meaning of the sentence that has been transformed, they also leave stylistic features of the text intact.

The relativiser switching method adopted in the text marking system developed by Murphy and Vogel (2007a) is very similar to the relativiser swap in the steganography system discussed above, but focuses more on irreversibility. In fact, none of the three methods proposed in this system is reliably reversible. The relativiser *who* and *which* preceded by a common noun is replaced with *that*. Unlike the relativiser swap in the previous system, *that* will not be changed into *who* or *which*, which will be a reversible process and also lower the security level as computers can still run into problems deciding whether the preceeding noun is human or non-human. The transformation is then verified by external structural fit, which compares the frequencies of the phrase containing the relativiser before and after transformation.

The adjective conjunction method switches the position of the two adjectives in the following pattern: "ADJECTIVE CONJUNCTION ADJECTIVE COMMON-NOUN". The potential transformation is then tested by internal structural fit and external structural fit. The internal structural fit compares the frequency of the ADJECTIVE CONJUNCTION ADJECTIVE

phrase before and after transformation, without considering the words preceeding and following the targeted phrase, while the external structural fit takes those words into consideration and compares the frequency of the left and right boundary word pairs.

The lexical substitution method uses corpus frequency data and graph topography to select individual content words in WordNet for lexical substitution. Their method does not use word sense disambiguation. It takes synonym sets of all the senses of the targeted word into consideration, then finds the synonym that maximizes the probability of the target word as a substitute. Another method proposed is to adopt balanced frequency substitution instead of maximized frequency substitution. Not only synonymy, but hyponymy and hypernymy are also exploited. The substitution is evaluated with the external structural fit and the semantic fit. The external structural fit functions in the same manner described in the other two methods discussed above, and the semantic fit tests the relative frequency of sets of surrounding content words, using word roots only and not considering the order of these words.

Their approach also differs from other system in the following ways: they do not use hand-craft lexical substitution sets; the approach does not need a word-sense disambiguation system, which would be a deeper kind of analysis and lacks accuracy.

2.7 Meral et. al.'s Morphosyntactic Transformation-Based System (2007, 2009)

Meral et. al.'s system (2009; 2007) is a morphosyntactic transformation-based linguistic watermarking system. The languages targeted are agglutinative languages, represented by Turkish, which are rich in morphosyntactic structure. But due to the fact that English is also included in their scheme and English examples are also given, this research nonetheless merits inclusion in the current study. However, for the English language, the methods that can be applied can be grouped syntactically. As their system was originally designed for watermarking rather than steganography, it focuses more on robustness, but the methods are quite transferable.

The following syntactic tools applicable to both Turkish and English are identified by Meral et. al. (2007) for developing their watermarking system, with active/passive voice, adverb displacement, and conjunct order change being the most frequently occurring (Fig 4).

Tools	Alternative 1	Alternative 2
Active/passive voice	Workers carried the sand.	The sand was carried by workers.
Adverb displacement	Ali will go to Istanbul tomorrow.	Tomorrow Ali will go to Istanbul.
Conjunct order change	Ali and Ayşe	Ayşe and Ali
Sentence 1: subject-predicate-DIR / Sentence 2: predicate subject-DIR	Watermarking is one of the important areas of study.	One of the important areas of study is watermarking
Subject-GEN verb- NOUN-POSS obvious / Obvious that subject-NOM verb-TENSE-AGR	That this work will not be so easy is obvious.	It is obvious that this work will not be so easy.

Fig 4. (Meral et al., 2007). edited)

As can be seen, the above transformations do not change the meaning and the stylistic features of a sentence. A customised corpus is used in their research. The input sentences taken from the corpus is in Treebank format, then transformed by the parser they developed into a hierarchical syntactical tree structure, which can mark the case of each token in the sentence and show the functional dependencies of each phrase and clause. The parsed and annotated Treebank sentence is then transformed back into an alternative syntactic tree representation. More than one transformation may be applied to each sentence.

2.8 Chang (2013)

In her PhD thesis, Chang (2013) proposed three possible linguistic transformations for steganography: lexical substitution, adjective deletion, and word ordering. Due to the encoding methods she adopts for her systems, a cover text is not needed for the receiver to decode the secret message in the stego text.

It should be pointed out that although in some cases the meaning of the cover text can be preserved after the text has been transformed by Chang's systems, especially for the lexical substitution-based system, meaning preservation and coherence at document level are not what her systems try to achieve. What her systems in fact strive to obtain is a natural stego

text, and all her proposed transformation checkers are evaluated by humans for sentence naturalness, which is by far the best way in terms of word-sense disambiguation.

2.8.1 Lexical Substitution-Based System

The lexical substitution method Chang proposed is synonym-based and similar to Bolshakov's system. Since in Chang's system, the message receiver will not have access to the original cover text, Chang also applies transitive closure to overlapping synonym sets in order to avoid decoding ambiguity. However, words that appear in the same synonym set as the cover word but which are not actually synonyms of the cover word are, for the reason mentioned above, only used for the encoding process and will not be used as candidates for substitution. Only synonyms of the cover word are considered candidates for the substitution, thus the possibility of unnatural substitution is lowered. This is one of the biggest differences between Bolshakov's system and Chang's system. The two systems also use different encoding methods. Chang's system uses synonym graphs together with vertex colouring to assign codes to words, unless they all belong to the same synonym set. In the latter case the block coding method will be applied. Another difference is that Chang's system does not use multiword. This will lower the number of candidates in each synonym set, but makes collocation verification easier.

Like Bolshakov's, Chang also constructs a checker to verify the contextual naturalness of the substitution based on the frequency of different collocations. Frequency counts of contextual bi- to five-grams of each of the synonyms in the synonym set are acquired from the Google n-gram corpus, which is reported to achieve better accuracy, precision, and recall values then using fewer n-grams. Words whose frequency score is lower than a certain threshold are discarded. In order to avoid high-frequency n-grams from dominating the substitution, the α-skew divergence is adopted to measure the difference between the n-gram count distribution of the most likely word and that of the candidate.

2.8.2 Adjective Deletion-Based System

Chang's adjective deletion system does not attempt to preserve meaning or achieve coherence at document-level, but does try to ensure that the sentences transformed are grammatically correct and semantically meaningful. Three precautions are taken to ensure the naturalness of the text. Firstly, a syntactic filter is used to perform a simple check of the grammatical correctness of the transformed sentence. Secondly, Google n-gram corpus is

used to check the fluency of the remaining context after the deletion. Thirdly, a support vector machine (SVM) model is trained to check if an adjective can be deleted in a given context.

The syntactic filter functions by first assigning a supertag, which is a lexical category expressing subcategorisation information, to all the words in the sentence apart from the targeted adjective, then checking if any of these supertags will be changed once the targeted adjective is deleted. If any of the supertags are changed, then the transformation does not pass the test and the adjective is considered not deletable.

Similarly, the checker based on Google n-gram corpus compares the bi- to five-gram counts before and after a deletion, and if the difference exceeds a certain threshold, the deletion is deemed unsuitable.

The support vector machine (SVM) model uses a hyperplane to classify adjective deletion cases. The SVM in this system uses contextual n-gram counts found in the Google n-gram corpus, two lexical association measures, adjective-noun modification entropies, and contextual α-skew divergence to determine the degree of association between an adjective and a noun. The two lexical association measures involved are Pointwise Mutual Information and log likelihood ratio, with the latter being able to handle rare events better than the former. The adjective-noun modification entropies are used to predict how unpredictable or fixed the modifier of a given noun is from the frequency of the different modification situations of that noun, while the α-skew divergence checks the divergence in various n-gram distributions of a noun before and after the adjective preceding it is deleted. If the n-gram distributions are similar before and after deletion, the targeted adjective is deemed deletable. As an obvious trade-off can be observed between the naturalness of the deletion and payload capacity, Chang allow the users of the stegosystems to adjust the threshold to achieve the balance they want.

The biggest drawback of Chang's adjective-deletion system is that it does not aim to preserve meaning or document-level coherence. To achieve preservation or near-preservation of meaning, one solution is to delete only redundant adjectives. While collecting pilot study data for developing this system, Chang only uses adjective-noun pleonasms such as *free gift*, *cold ice*, and *final end* as positive data for training the SVM, but due to considerations of payload capacity, the final product does in fact delete other adjectives. If meaning preservation is to

be achieved, the system can be limited to only deleting pleonasms. Pleonasms may also be explored further to include other kinds of pleonasms for deletion.

2.8.3 Word Ordering-Based System

Chang's stegosystems based on word-ordering are closer to linguistic generation-based stegosystems than linguistic transformation-based systems. Like the adjective-deletion stegosystems, Chang's focus is not on meaning preservation or document coherence but only how natural they appear.

The word-ordering stegosystem developed by Chang functions by dismantling a sentence into a set of un-ordered words: a bag-of-words, then reassembling them with a word ordering realisation system. Zhang et. al.'s (2012) system is chosen for the realisation system. Permutations that contain fewer words than the bag-of-words are eliminated to achieve a higher security level. Chang proposes two checkers to check the naturalness of a permutation: a baseline word ordering checker using Google n-gram corpus and a maximum entropy classifier. The Google n-gram baseline word checker, which compares the n-gram counts before and after word ordering, proves to be ineffective, probably due to the fact that the longest n-gram is five-gram, which may be not long enough for checking sentence-level performance. The maximum entropy classifier has much better performance. It is trained using some syntactic features to determine the naturalness of a sentence permutation so as to eliminate any unnatural permutations. Similarly to the adjective-deletion stegosystems, users can adjust the threshold to control the trade-off between precision and recall.

One drawback of this word-ordering stegosystem, especially when compared with Atallah et. al.'s (2003), is its low payload capacity. Chang's system exploits the LSB of every stego sentence to encode messages, with the embedding scheme's upper bound being 1 bit per sentence. In addition, there is the possibility that no permutation can match the bitstring desired.

Like Atallah et. al. (Atallah et al., 2001; Atallah et al., 2003), Chang also finds her word-ordering system related to machine translation. Firstly, both Atallah et. al. and Chang see that their system resembles a machine translation system that translates between the same language. Secondly, the BLEU metric, which was originally designed for automatic evaluation of machine translation, is found to correlate highly with the human judgement made on the sentences generated by Chang's word-ordering stegosystem. The hash function

encoding method adopted by Chang for this system is also the encoding method used in existing translation-based stegosystems, such as that of Grothoff et al (2005).

Although Chang claims that her system focuses on maintaining the naturalness in the stego text, as her system neither tries to preserve meaning nor attempts to achieve textual coherence, it is possible that human readers will still find the stego text "unnatural". However, many of her approaches can be useful in enlarging and checking existing techniques in meaning-preserving stegosystems.

2.9 Wang et. al. (2013)

Wang et. al.'s context-based substitution system uses the conditional equivalence between the definite article (the) and demonstrative adjectives (this, that, these, those) to embed secret message. They adopted a part-of-speech filter, a 3-gram language model, a 2-range context, the minimum matching distance, and the perplexity to evaluate the substitution. The system is reported to be highly accurate and reliable, but suffers from low payload capacity.

Dictionaries of noun, adjective, and adverbs are built using the most frequently used words in 500 popular English novels, which are also their training text set. Some editing is carried out manually. In the noun dictionary, all the abstract nouns are removed; single and plural forms of each noun are listed. All the possessives are removed from the adjective dictionary. Finally, each token can only appear in one dictionary.

Since *this*, *that*, *these*, *those* can have other functions other than demonstrative adjectives, a part-of-speech filter is adopted. Only if a demonstrative adjective or a definite article is followed by "NOUN", or "NOUN-NOUN", or "ADJECTIVE-NOUN" or "ADVERB-ADJECTIVE-NOUN" will it be accepted as potential candidates for substitution.

When a demonstrative adjective is preceeding an object, the demonstrative adjective can always be substituted by a *the*, thus, as long as a potential clause has passed the POS filter, the demonstrative adjective can be substituted by the definite article. However, if a definite article is to be replaced by a demonstrative adjective, the object following the definite should be mentioned before, unless an attributive clause is involved. Wang et. al. decided to leave out the attributive clauses for the complexity of the situation. As to judging if the object involved has been mentioned before, they decide to simply check for the existence of the word in the preceding, leaving out the situations where the same object is mentioned using different expression.

They adopt the Minimum Matching Distance (MMD) to check the distance between the targeted object and the nearest same one in the preceding. Results show that the MMD mainly falls in the closed interval from 0 to 14 sentences.

Another problem is brought by the phenomenon of zero articles in clauses led by *that*. To solve this issue, an m-range context is built to discover potential collocations in the conflict and remove them. A 3-gram model is used to estimate the probability. A word perplexity is applied to the 2-range context to tackle the collocation problem. The substitution suitability degree is adopted to evaluate the suitability of the substitution, considering MMD, which needs to be lower than 14, the PPL, which should not be too high.

For the evaluation process, a novel by Mark Twain is chosen, and 40 humans were invited to evaluate the stego text. The results show that this method has a low error rate and low quantity of unsuitable substitutions. When the embedding rate is set at 6.22 bit per KB, the error rate is 5.67%, when it is set to 4.76 bit per KB, the error rate drops to 1.32%. It takes a further drop to 0.3% when the embedding rate is 1.0 bit per KB.

This method has several advantages. Firstly, the transformation can easily preserve the meaning and style of the text, and is likely to be grammatical. Secondly, considering the prevalence of definite article and demonstrative adjectives in natural language text, it is quite likely that some potential places for substitution can be easily found. The authors also claim that they cannot discover any anomaly by the analysis of the occurrence frequency since these words are so common. Lastly, the error rate reported is low compared with some other linguistic steganography techniques.

One of the drawbacks of this system is its comparatively low embedding bit rate. It would be interesting to see how the payload capacity of this system might be improved by considering more potential possibilities for substitution that have been left out by the authors due to their complexity and how that may affect the security level of this system. It would also be beneficial to see how the system would function using a larger dictionary instead of limiting the vocabulary to the most popular 500 English novels, such as using WordNet. Why they decided to choose a 3-gram model is not fully explained, and it would be interesting to see if other n-gram models, or a combination of them, will lead to better performance.

2.10 Other Systems

Two other systems using linguistic transformation for steganography are introduced here. These two systems are less important as one does not propose new methods of linguistic transformation and the other is only a speculation.

2.10.1. Wai and Khine's Approach Using Syntax Bank and Digital Signature (2011)

Wai and Khine's (2011) main contribution is not in the linguistic transformation method they adopted for their steganography system, but in the way they encrypt the message. However, since the method involved is meaning-preserving and linguistic transformation-based, it is included in this paper.

A syntax bank in their system is made up of a number of syntax sets, which are shared between the sender and the receiver. A syntax set contains all available syntax forms of a sentence. In their system, each syntax form in a syntax set is semi-randomly assigned a binary number.

During the encoding process, a sentence is parsed, while at the same time the secret message is compressed by Shannon-fano algorithm. The parsed sentence is then transformed into one of the syntax forms within the syntax set of the cover sentence. In the last step digital signature of the stego text is generated using SHA-512 hash algorithm, digital signature algorithm, and the sender's private key.

During the linguistic transformation process, once the system decides that the input sentence's syntax is available, the syntax is divided clause by clause. The syntax of the first clause is searched for in the syntax bank, and once found, all the syntax forms with the same set number will be extracted. The current system only looks at active-passive transformation.

The system is reported to have an average payload of 0.6 bit per sentence. 20 human beings are invited to participate in evaluation, and a success rate of 95% is reported.

2.10.2 Kamel & Banawan (2012)

Kamel and Banawan (2012) proposed a meaning-preserving syntactic-based method for linguistic steganography. Their scheme is based on relocating maneuverable sentence constituents.

Their system identifies candidates by spotting specific keywords, such as *and*, *separated*, and *unseparated*. They have identified four transformation types: splitting separable phrasal verbs, such as *fill up, try on*; swapping operands of a Boolean operator, such as operands around *and*, *or*, and *nor*; re-arranging conditional statement, such as clauses with *if*, or *because*; and adverb replacement. Here are some examples taken from Kamel and Banawan (2012).

A. Splitting separable phrasal verbs.
   Sentence 1: *Take off those bright yellow sandals.*
   Sentence 2: *Take those bright yellow sandals off.*
B. Swapping operands of Boolean operator
   Sentence 1: *He enjoys reading both magazines and novels.*
   Sentence 2: *He enjoys reading both novels and magazines.*
C. Re-arranging conditional statement
   Sentence 1: *If you pay using cash, you will get a discount.*
   Sentence 2: *You will get a discount, if you pay using cash.*
D. Adverb placement
   Sentence 1: *Also, we can start on the research project tonight.*
   Sentence 2: *We, also, can start on the research project tonight.*

Kamel and Banawan (2012)

Notice that the sample given for the transformation type "adverb placement" can has other variations.

However, their system is still just a speculation. They did not provide the algorithms for implementing these tasks. They claimed the method to be meaning- and style-preserving, but did not supply any test results. They claim the embedding capacity to be higher than 0.8 bits per sentence, yet the evaluation is done solely by carrying a manual search on Google Books Corpus. The encoding method adopted is block coding and rather simple.

Chapter 3

Discussion

3.1 On Meaning Preservation

As steganography functions by taking advantages of the noises in the cover datatype, it is important to find out what can be the noise in natural language. "The task, then, is in finding degrees of freedom in some textual entity that allow an underlying work to be changed while preserving its meaning." (Keith Winstein, 1998)

From the discussion in the previous chapter, it can be seen that meaning is the key factor in future research. Firstly, it is very important for security level. Secondly, at the current state, computers are better at syntactic analysis then meaning analysis. From Winstein to Bolshakov to Atallah etc., many designers of steganography systems are striving to achieve better word-sense-disambiguation.

However, is it really possible to preserve meaning while carrying out any linguistic transformation? When something is deleted from a sentence, even just an adjective, unless it is purely redundant information (perhaps due to our "fact database" or some presupposition as some people might argue), or when an adjective is substituted, some nuanced difference in meaning can be introduced. When some elements are added, or when a clause is changed, it can also result in some stylistic differences, or bringing a certain meaning to the front ground, shifting our attention.

There are certainly different levels of meaning preservation, such as sentence-level meaning preservation, paragraph-level meaning preservation, and document-level meaning preservation. Syntax-based methods excel in meaning preservation compared with many other methods, especially sentence-level meaning preservation. For those developing methods based on semantic transformation and deeper techniques, meaning preservation can be understood as existing on a higher, more contextual level.

It can also be seen that the state-of-the-art of word-sense-disambiguation is still very immature, thus seriously impeding steganography related to meaning processing, especially systems based on lexical substitution. That also means that we still need to use humans for an effective evaluation of any linguistic stegosystem. At the moment, there is a lack of benchmarks and methodology for objectively evaluating a linguistic stegosystem, and

different researchers often use very different ways to evaluate their results, making it difficult to compare.

3.2 Shallow vs. Deep

The dilemma between using shallow or deep techniques, can also be reflected as the dilemma between syntax-based and semantics-based techniques.

The unsatisfactory state of computer word-sense-disambiguation also means relying on shallow parsing techniques and syntax-based approach in general will have a much better result in accuracy and reliability compared with techniques using deeper techniques such as deep parsing and word-sense-disambiguation, as is proven by the many examples analysed in the previous chapter. Considering the current state-of-the-art of NLP techniques, building a system based on shallow and syntactic approaches are much more likely to succeed than relying on WSD or deep parsing.

However, despite the current problem in WSD and deep parsing, from a purely theoretical point of view, methods based on semantic transformation do promise a better potential for increasing payload then methods based on syntactic transformation, and the ontological semantics theory does look tempting for the future, especially when we consider how little noise languages contain compared with other media such as video and audio, leading to its low payload capacity. However, considering computers' poor ability in understanding meaning, any kind of insertion or generation based methods can be dangerous.

It is interesting that more than one scholar mentioned in the previous chapter sees linguistic steganography system or watermarking system as a special form of machine translation, and suggest using machine translation evaluation method to evaluate a stegosystem or watermarking system. These scholars tend to adopt the approach of turning surface structure into deep structure, and then converting the deep structure back into the surface structure, be it a syntactical structure or a semantic structure, or both. If such an approach is taken, then indeed, both linguistic steganography and machine translation are trying to turn a surface structure, both syntactic structure and semantic structure, into a deeper structure, and then regenerate an alternative surface structure from the deep structure.

Different languages have different features. Some languages naturally work better than other languages when it comes to some specific methods, as can be seen by the morphosyntactic-based methods proposed by Meral et. al. (2009), which work better with Turkish than English.

Yet compared with semantically related methods, it is likely that syntactically related methods are more portable then semantic ones.

3.3 Ontology vs. Statistics

In the stegosystems studied in the previous chapter, they either use some ontology or structural approaches, such as compiling meaning-defining dictionaries, or statistical approaches, represented by n-gram approach. They are usually used together to achieve a better result.

Very often, ontology is applied first to a system, then statistical methods are introduced to verify or limit the data to achieve better results. If ontology can be understood as a prescriptive approach, statistical methods represent a descriptive approach.

As Chang (2013) has mentioned in the future work section of her PhD thesis, automatic word insertion as a potential way of transformation may create more alternatives for a cover text, increasing the payload capacity. Given the state-of-the-art of NLP, this will be difficult, but Atallah et. al (2003) have also shown us that it might not be totally impossible in the future. One of the difficulties lies in the "ontology" of the world, which ontological semantics is trying to solve. Weather ontological semantics is really useful, or to what extent it is useful, is still a question, but it may be helpful to use statistical methods, represented by n-gram counts, combined with machine learning, like the maximum entropy classifier and SVM, to be a checker to verify the ontology-based database. It will be even better if it is possible to automatically update this database, but that, paradoxically, will require better NLP techniques.

The difficulty with ontology lies in the fact that it needs to be handcrafted and is difficult to scale up. Besides, it is not very portable. When a different language is involved, a new ontology needs to be written. Comparatively, statistical methods are much easier to be transferred.

Conclusion

This research paper has looked at some basic concepts in steganography and linguistic steganography, introduced and discussed the different stegosystems based on linguistic transformation, including systems based on lexical substitution or deletion, word ordering, syntactic transformation, and semantic transformation. It has also tried to identify some basic and general patterns in these different systems.

Several conclusions are drawn: meaning preservation is a key in linguistic steganography, and an important future research direction, yet there are many different levels of meaning preservation. Humans are absolutely necessary in evaluating how well a steganography system preserves meaning. A common evaluation framework or benchmarks are needed for evaluating linguistic steganography systems. Stegosystem relying on shallow techniques are more reliable an accurate at the moment but semantic-based methods should be able to provide us with higher payload capacity if NLP techniques are more advanced. As machine translation and linguistic steganography based on semantic transformation are similar in some ways, the advance in one is likely to benefit the other. The final conclusion is that some techniques or methods may be more portable or scalable then others, which can be taken into consideration if a scalable and portable system is to be developed.

Due to time constraint, I do not have the chance to put all the systems discussed into perspective and to do a thorough comparison and contrast of these systems. It would be beneficial if all the stegosystems can be categorised according to the methods they are based on, and then carry out a comparison within each category. Another problem with this research is that different systems provide different evaluation methods and parameters, making it difficult to compare. It would be much clearer if a table could be made, with all the different parameters and evaluation results listed. An issue that can be looked into is now different stegosystems perform when dealing with texts within different genre and registers. It may also be helpful to perhaps look at machine translation, as it can be related to some of the approaches adopted in linguistic steganography. This survey only deals with stegosystems targeting the English language, looking at how stegosystems written for other languages and how different systems may be adapted for different languages may also lead to some interesting observations.

References

Atallah, M. J., Raskin, V., Crogan, M., Hempelmann, C., Kerschbaum, F., Mohamed, D., & Naik, S. (2001). *Natural language watermarking: Design, analysis, and a proof-of-concept implementation.* Paper presented at the Information Hiding.

Atallah, M. J., Raskin, V., Hempelmann, C. F., Karahan, M., Sion, R., Topkara, U., & Triezenberg, K. E. (2003). *Natural language watermarking and tamperproofing.* Paper presented at the Information hiding.

Bennett, K. (2004). Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text.

Bergmair, R. (2004). Towards linguistic steganography: A systematic investigation of approaches, systems, and issues. *Final year project, The University of Derby, April*.

Bergmair, R. (2007). *A comprehensive bibliography of linguistic steganography.* Paper presented at the Electronic Imaging 2007.

Bolshakov, I. A. (2005). *A method of linguistic steganography based on collocationally-verified synonymy.* Paper presented at the Information Hiding.

Bolshakov, I. A., & Gelbukh, A. (2004). Synonymous paraphrasing using wordnet and internet *Natural Language Processing and Information Systems* (pp. 312-323): Springer.

Chang, C.-Y. (2013). *Transformations for linguistic steganography.* (Doctor of Philosophy), University of Cambridge, Cambridge.

Chapman, M., & Davida, G. (1997). Hiding the hidden: A software system for concealing ciphertext as innocuous text. *Information and Communications Security*, 335-345.

Chapman, M., Davida, G. I., & Rennhard, M. (2001). A practical and effective approach to large-scale automated linguistic steganography *Information Security* (pp. 156-165): Springer.

Cox, I. J., & Miller, M. L. (1997). *Review of watermarking and the importance of perceptual modeling.* Paper presented at the Electronic Imaging'97.

Desoky, A. (2009a). Listega: list-based steganography methodology. *International Journal of Information Security, 8*(4), 247-261.

Desoky, A. (2009b). Notestega: notes-based steganography methodology. *Information Security Journal: A Global Perspective, 18*(4), 178-193.

Desoky, A. (2011). Edustega: an education-centric steganography methodology. *International Journal of Security and Networks, 6*(2), 153-173.

Fellbaum, C. (1999). *WordNet*: Wiley Online Library.

Fridrich, J. (2009). Steganography in Digital Media: Principles. *Algorithms, and Applications, Cambridge, England: Cambridge University Press*.

Grothoff, C., Grothoff, K., Alkhutova, L., Stutsman, R., & Atallah, M. (2005). *Translation-based steganography.* Paper presented at the Information Hiding.

Johnson, N. F., & Katzenbeisser, S. (2000). *A survey of steganographic techniques.* Paper presented at the Information hiding.

Kahn, D. (1996). *The History of Steganography*. Paper presented at the Proceedings of the First International Workshop on Information Hiding.

Kamel, I., & Banawan, S. (2012). *Hiding information in the placement of maneuverable words.* Paper presented at the Innovations in Information Technology (IIT), 2012 International Conference on.

McCarthy, J. (1990). An example for natural language understanding and the AI problems it raises. *Formalizing Common Sense: Papers by John McCarthy. Ablex Publishing Corporation, 355*.

Meral, H. M., Sankur, B., Sumru Özsoy, A., Güngör, T., & Sevinç, E. (2009). Natural language watermarking via morphosyntactic alterations. *Computer Speech & Language, 23*(1), 107-125.

Meral, H. M., Sevinc, E., Ünkar, E., Sankur, B., Özsoy, A. S., & Güngör, T. (2007). *Syntactic tools for text watermarking.* Paper presented at the Electronic Imaging 2007.

Murphy, B., & Vogel, C. (2007a). *Statistically-constrained shallow text marking: techniques, evaluation paradigm and results.* Paper presented at the Electronic Imaging 2007.

Murphy, B., & Vogel, C. (2007b). *The syntax of concealment: reliable methods for plain text information hiding.* Paper presented at the Electronic Imaging 2007.

Nirenburg, S., & Raskin, V. (2001). *Ontological semantics, formal ontology, and ambiguity.* Paper presented at the Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001.

Nirenburg, S., & Raskin, V. (2004). *Ontological semantics*: MIT Press.

Petitcolas, F., Anderson, R., & Kuhn, M. (1999). Information hiding - A survey. *Proceedings of the Ieee, 87*(7), 1062-1078. doi: 10.1109/5.771065

Pfitzmann, B. (1996). *Information Hiding Terminology - Results of an Informal Plenary Meeting and Additional Proposals*. Paper presented at the Information Hiding.

Shih, F. Y. (2007). *Digital watermarking and steganography: fundamentals and techniques*: CRC Press, Inc.

Simmons, G. J. (1984). *The prisoners' problem and the subliminal channel.* Paper presented at the Advances in Cryptology.

Spammimic.com. Retrieved Feb 20, 2014, from http://www.spammimic.com/

Topkara, M., Taskiran, C. M., & Delp III, E. J. (2005). *Natural language watermarking.* Paper presented at the Electronic Imaging 2005.

Topkara, M., Topkara, U., & Atallah, M. J. (2006). *Words are not enough: sentence level natural language watermarking.* Paper presented at the Proceedings of the 4th ACM international workshop on Contents protection and security.

Topkara, U., Topkara, M., & Atallah, M. J. (2006). *The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions.* Paper presented at the Proceedings of the 8th workshop on Multimedia and security.

Vybornova, O., & Macq, B. (2007). *Natural Language Watermarking and Robust Hashing Based on Presuppositional Analysis.* Paper presented at the IRI.

Wai, E. N. C., & Khine, M. A. (2011). *Syntactic Bank-based Linguistic Steganography Approach.* Paper presented at the 2011 International Conference on Information Communication and Management IPCSIT.

Wang, F., Huang, L., Chen, Z., Yang, W., & Miao, H. (2013). *A novel text steganography by context-based equivalent substitution.* Paper presented at the Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on.

Wayner, P. (1992). Mimic functions. *Cryptologia, 16*(3), 193-214.

Wayner, P. (2009). *Disappearing cryptography: information hiding: steganography & watermarking*: Morgan Kaufmann.

Winstein, K. (1998). Lexical steganography through adaptive modulation of the word choice hash. *Unpublished. http://www. imsa. edu/~ keithw/tlex*.

Winstein, K. (1999). Tyrannosaurus lex. *Website, January*.

Zhang, Y., Blackwood, G., & Clark, S. (2012). *Syntax-based word ordering incorporating a large-scale language model.* Paper presented at the Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics.