情報学基礎 57−13
自然言語処理 136−13
(2000. 3. 22)

意味保存型の情報ハイディング
—日本語文書への適用—

三瓶 光司* 中川 裕志† 松本 勉* 村瀬 一朗‡
横浜国立大学 工学部* 東京大学 情報基盤センター† 三菱総合研究所‡

ネットワーク基盤の発達等により，電子的にやり取りされる情報量が目覚ましく増加している．それに伴い，電子化されたコンテンツに対する著作権保護が大きな問題となっている．その問題を解決する一つの方法として注目，研究されているのが情報ハイディングである．しかし，従来の情報ハイディングは，画像，音声などに対するものがほとんどであり，テキストを対象にしたものでも，文字間や行間などを微妙に変化させて情報の隠蔽を行うなど，実質的には画像的な扱いをするものがほとんどであった．そこで我々は，自然言語処理の技術を情報ハイディングの分野に応用し，文章中の表現を意味の変わらない他の表現に置き換えることで情報隠蔽を行うシステムを提案する．本稿では置き換えに敵した言語要素の考察，置き換え規則およびその記述，システム構成，さらにソフトウェア附属文書を対象し，置き換えによって情報を隠蔽したテキストの自然さの評価などについて報告する．

Meaning Preserving Information Hiding.
—Japanese Text Case—

Koji SAMPEI* , **Hiroshi NAKAGAWA†** , **Tsutomu MATSUMOTO***
and **Ichiro MURASE‡**
Faculty of Engineering, Yokohama National University*
Information Technology Center, the University of Tokyo†
MITSUBISHI Research Institute Inc.‡

Information hiding methods becomes paid growing attention these days. Many systems of information hiding are proposed as one of the technique resolving copyright problems in digital contents. In fact, almost all of existing information hiding methods are intended for images and sounds. Methods of hiding information into text proposed so far, however, processed texts as essentially an image. In this paper, we propose an information hiding method for text which uses text as not paper images but sequences of character symbols. The means that linguistic expressions are altered to hide hidden information while the meaning of the text is still preserved. Our target text is written in Japanese in the field of software manual and document for user agreement of the software. Actually, this method hides information into text by paraphrasing with a dictionary which consists of pairs of expressions having the same meaning. We develop two types of dictionary. One is for general linguistic expressions, and the other is including domain specific terminologies' paraphrasing expressions, too. We report here the experimental evaluation our method.

1 はじめに

ネットワークの発達や大容量記憶メディアの低価格化に伴い，電子化された画像，音声，ドキュメント等の流通量が著しく増加している．このようなデータは，複製，加工等の操作が容易であるため，著作権の保護が大きな問題となっている．以前はこれらの問題に対処する手段として，データ転送時の安全性と機密性に優れる暗号化技術が利用されてきた．しかし，暗号化技術では受信側

で復号されたデータに対して，複製，加工等を阻止することは完全には不可能であった．

そこで，これら違法な複製等を抑制する手段として注目，研究されているのが情報ハイディングである．情報ハイディングは，元データの品質を落とすことなくそこに情報を隠蔽しておき，必要に応じてそれを検出する技術である．この情報ハイディングは，主として画像，音声データに対する分野で研究がなされ実用化が進んでいる．一方，テキストデータに対するものは事例が少なく，現在有効な手法が模索，研究されている．なお，暗号技術と比較すると，暗号は情報の内容を秘匿する技術であり，情報ハイディングは情報の存在そのものを隠す技術であると位置づけられる．

本稿では，我々が開発した，ある言語表現を意味を変えない別の表現で置換する情報ハイディングを提案する．また，今回開発したシステムの評価についても報告する．

2 テキストへの情報ハイディング

2.1 情報ハイディングシステムの枠組み

情報ハイディング技術は電子透かし，電子指紋等の基盤となる技術であり，ある対象に対し，秘匿情報を埋め込む技術である．一般に，秘匿情報が埋め込まれる対象となるデータをカバーデータ，カバーデータに情報を秘匿することによって生成されたデータをステゴデータ (stego data) と呼ぶ．本稿では，テキストへの情報ハイディングについて論ずるので，カバーデータ，ステゴデータは，それぞれ，カバーテキスト，ステゴテキストとなる．

2.2 従来の方式

テキストへの情報ハイディング技術は，大きく分けて 3 種類の方式に分類できる．以下に，各方式の特徴を示す．

1. ホワイトスペース法 [1, 2]
   画像に近い形式のテキストデータにおいて，行間，文字間を微妙に変化させることにより情報を隠蔽する．
   文章の内容には手を加えないため，秘匿情報の存在を見破られにくいが，文面だけをコピーされたり，変形をされた場合，何の効力も持たなくなる．

2. 文字埋め込み法 [6, 7]
   スペースやタブ等を単語間や行末等に埋め込むことにより情報を隠蔽する．
   文書に不自然なスペース，タブ等が挿入されるため，秘匿情報の存在を見破られやすく，ホワイトスペース法と同様の欠点を持つ．

3. 辞書変換法 [3, 4, 5, 8]
   文法構造と，各単語が持つ秘匿情報を保持した辞書をあらかじめ用意し，その辞書を用いて秘匿情報に合わせた文章を構築する．
   文法構造のみを考慮するため，生成された文書に意味のつながりはなく，不自然なものとなるため情報の存在を見破られやすい．

従来の方式では，テキストを画像データとして処理するものが多く，また，テキストデータとして処理するものでも，文法構造のみを考慮し，意味は保存されないものであった．そこで我々は，意味情報を保存するテキストへの情報ハイディングを提案する．

3 意味保存の情報ハイディング

3.1 従来方式の拡張

我々は，同一の内容であっても複数の表現が存在することに着目した．例えば，「意味を変えない置き換え」という文に対しては，「意味を変化させない置き換え」，「意味を変えない置換」，「意味を変化させない置換」といった，同一の意味を持つ文が存在する．仮に，「変えない」と「置き換え」がビット情報 0 に，「変化させない」と「置換」がビット情報 1 に対応しているとすれば，これら意味の同一な 4 つの文はそれぞれ，“00”，“10”，“01”，“11” の情報を表わしていることになる．このような表現の置換を文書全体に対して行えば，文書の品質が損なわれず，かつ，秘匿情報の存在を検出されないステゴテキストが生成可能であると考えられる．

3.2 表現の置換

3.2.1 一般的な置換

我々が普段使っている言語には，同一の内容を表現するものであっても，多種多様な表現が存在する．前にあげた，「意味を変えない置き換え」のような，各語を同一の意味を持つ語で置き換えるようなものもあれば，語順を変えるもの，全く違う構文，語で同一の内容を表現するものなど様々である．我々のシステムの対象となるものは，情報ハイディングという分野から，主として文書となるため，ここでは文書における表現の置き換えについて考察する．

A. 句の順番入れ替え

(1)「船で沖縄に行く」のような，1つの語に対して2つ，またはそれ以上の語がかかる場合，(2)「暗号とセキュリティ」のような，2つ以上の語が並列句をなす場合など，は語順を入れ換えられる可能性がある．単語を単独に扱える場合なら入れ換えは簡単だが，実際の文書においては，関係節などの修飾部分を伴う句が並列である場合が多い．このため計算機においてこれらの語順の置換を行う場合には，関係節の範囲，係り受け構造の厳密な認識が必要となるため，正確な構文解析システムが必要となる．また，2つ以上の語や句が等価である場合については，それらの表現が用いられた後，前者，後者などといった参照の仕方をすることがある．その場合には構文解析に加え，意味解析も必要となる．

B. 受動態から能動態へ，能動態から受動態へ
例えば,「システムが構文を解析する」であれば,「システムによって構文が解析される」となる．しかし，先に述べた語順の入れ替えと同様，句を構成している場合にも正確に変換するためには，正確な構文解析が必要である．

C. 同義語，類義語の利用
文中の語または語句を，それと同一の意味をもつ別の表現で置き換える．文中の一 部を置換するため，構文解析は必要としないが，置き換える語の活用形などで同義語が変化することもあるため，それらの活用形などを調べる必要がある．また，同義語は各々の語が持つ意味だけに依存するため，意味が同一な語や語句の対応を記述したデータベースを予め用意しておく必要がある．

D. 送り仮名，仮名⇔漢字等の表記揺れを利用する
「表す，表わす,」「ください，下さい」等がその例であるが，これらもあらかじめデータベースを作成しておかないと置換することはできない．また，これらはあらかじめ文書全体で統一されていることが多いため，置き換えによって，意味は変化しなかったとしても，文書全体の統一性が崩れ文書が不自然なものとなることがある．

E. 冗長である部分を付加，削除する
「して，し,」「しなければ，せねば,」等，主に平仮名を付加または削除することで置き換えができる．同義語や表記の揺れを用いたものと同様，データベースの作成が必要である．一般に，冗長である部分を削除する方が周囲に与える影響は小さい．付加すると，余分なものがつくわけであるから，文が回りくどい表現になってしまうことがある．

上記のもののうち，A. と B. は構文解析または意味解析が必要となってくる．これらの処理は，現在の自然言語処理技術では曖昧な結果しか得られないことが多く，実用性に問題がある．また，これら構文解析を用いる置き換えの手法は，文節や文を単位とするので，同義語等を用いた場合と比べ隠蔽できる情報量が少ない．以上の考察により，我々のシステムでは，文の構造を変えるのではなく，C. D. E. に述べた語や語句の置き換えによって情報を隠蔽する方式をとることにする．

3.2.2 置換条件の設定

表現の置換において，最も考慮しなければならないことは，意味構造を崩さないこと，文法構造を崩さないことである．これを実現するために，形態素解析システム茶筌 [9] によって品詞タグ付けされた日本語テキストにおいて，置換可能な際の条件をあらかじめ設定する．

一般に，置き換えられた表現に近い語ほど，置き換えによって受ける影響が大きい．我々はこのことを考慮し，置換の対象となる表現の周囲の語によって置換可能かどうかの判定を行うような条件を設けた．形態素解析によって得られる情報は，品詞の種別やその活用形であるため，条件を構成する要素として，置換の対象となる語からの距離 (形態素数) と，語そのもの，品詞名を選択した．条件は論理式に近い形式で記述され，辞書に登録される．条件は基本的に，

「位置特定子」+「 == または!= 」
+「品詞識別子または文字列」

のように記述される．位置特定子とは図1に表わしたものであり，置き換え可能かどうかを判定される語を Self，それ以前に出現する語を Self から近い順に Pre1，Pre2，…，Pre5，Self 以降に出現する語を近い順に Post1，Post2，…，Post5 といった名前をつけたものである．また，品詞識別子とは，名詞，動詞，形容詞等にそれぞれ別名をつけたものであり，これらであれば，Noun，Verb，Adj という風になる．また，複数の条件を ||，&&でつなげることもできる．例えば「直前が名詞でなく，直後が動詞でないならば置き換え可能」という条件を記述すると，(Pre1!=Noun)&&(Post1!=Verb) となる．

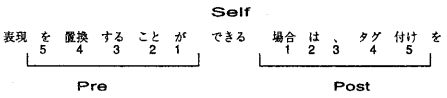

図 1: 位置特定子

3.2.3 品詞の種類による特徴

置換の対象となる品詞の種類によって，同義語への置き換えは，さらに細かく分類することができる．例えば，普通名詞であれば置き換えの際，

語の活用について考慮する必要はないが，複合名詞化しているものは，単純に置き換えると文法構造が崩れる恐れがある．また，動詞は活用形を考慮して置き換えねばならない．これら品詞の種別による表現の置き換えの特徴を示す．

○ 名詞 (サ変名詞[1]を除く)
前方および後方に名詞が連接する場合を除き，文法的制限はない場合が多い．しかし意味的に完全に等しい語が存在する場合は少なく，特定の条件下において同じ意味を持つ場合や，一方がもう一方を包含する意味を持つことが多いため，一方向の変換になることが多い．「使用」と「利用」等がその例である．また，一般的に用いられる名詞であっても，特定の分野においては特殊な意味を持つ場合があるため注意が必要である．

○ サ変名詞
サ変名詞単体で文中に出現する場合は普通の名詞と同様の特徴をもつが，「サ変名詞＋する」の場合には動詞として振る舞うため，この場合は動詞の置換と同様，「する」の部分の活用形を考慮した置換を行わねばならない．また，「～を＋サ変名詞＋する＋こと」のようにサ変名詞を含む名詞句が存在する場合は，この部分を「～の＋サ変名詞」とする置き換えが可能である．例えば「表現を置換することは，」であれば，「表現の置換は，」となる．

○ 動詞・助動詞
動詞および助動詞の置換に当たってもっとも考慮しなければならないことは，活用形の変化である．例えば，「できる」という助動詞とそれに対する置換候補「可能だ[2]」を考えた場合，前者は終止形，連体形とも「できる」であるが，後者は終止形が「可能だ」，連体形が「可能な」となる．このように置換対象とその置換候補では必ずしも活用形が一致するわけではないので注意が必要である．
また，サ変名詞が名詞句を形成する場合と同様，「～を＋動詞＋こと」は，これと同様の意味を持つ名詞を用いて，「～の＋名詞」とすることが可能である．「使うこと」と「使用」であれば，「計算機を使うことは，」が「計算機の使用は，」となる．

○ 形容詞
形容詞にはイ形容詞，ナ形容詞の 2 種類がある．基本的に形容詞は同一の意味を表す語が少なく，置き換えの対象となる語，およびそれに対する候補によって，考慮せねばならないことが変化する．イ形容詞であれば，同一の意味を持つ表現はほぼイ形容詞に限定されるため文法構造は考慮する必要がない場合が多いが，各語ごとの微妙な意味合いの違いを考慮しなければならない．また，ナ形容詞の場合は，ナ形容詞自身の活用を考慮した置き換えを行わなければならない．

○ 副詞，接続詞
副詞，接続詞は同一の品詞同士の置換が可能なことが多く，また，活用もしないため，単純に置換できるものが多数存在する．しかし，置換対象とその候補を文語体と口語体のような組み合わせにしてしまうと文書全体の統一性が失われる恐れがある．例えば，「のみ」と「だけ」がその様な語にあたる．

以下に分野によらず適用可能な置き換え表現の集合 (以下，これを一般辞書と呼ぶ) の一部分を示す．

```
(関連した, 関連する,
   (Post1==Noun)&&(Post==EndOfSentence))
(有する, もつ, Null)
(できる, 可能な,
   (Post1==Noun)&&(Post1==EndOfSentence))
(できる, 可能である, Null)
(別の, 他の, Pre1==EndOfSentence)
(⟨サ変名詞⟩して、, ⟨サ変名詞⟩し、, )
```

図 2: 一般辞書の例

3.3 システム概要

本システムの処理の流れを以下に述べる．

- カバーテキスト C を形態素解析によって各形態素に分解する．
- 文書変換辞書 D の構造は図 3に示すように，置換対象となる語と，それに対する置換候補，置換の際の条件を保持する．
- 形態素解析されたテキスト C と，辞書 D の登録内容との比較を行い，D 中に存在する表現があれば，置換条件による判定を行う．置き換えが可能であった場合は，置き換えが可能である語と共に，それに対する候補もタグ付けがなされテキスト C' として出力される．
- テキスト C' は HTML 形式に変換する専用のフィルタをかけることで，どのように置き換えられるかが Web ブラウザで確認できるようになっており，C' 中の置き換えに不適切な部分があった場合，人手でそれを修正することが可能である．
- 秘匿情報 e はバイナリストリングに変換され，そのバイナリストリングと C' のタグ情報にしたがって表現の置換が行われる．その結果生成されるのがステゴテキスト S である．
- 秘匿情報の抽出はテキスト C' とテキスト S の比較によって行われる．比較によってどの個所が置き換わっているかを特定し，情報隠

[1]後方に “する” が続くことで動詞を形成する名詞．“置換” など．

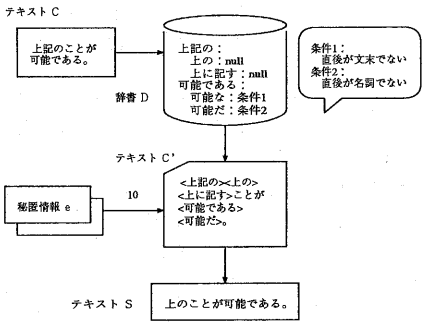

図 3: 辞書の構造

蔽の際に適応した情報埋め込みパターンを用いて情報の抽出を行う．

上記のような方法で作られたステゴテキストの一例を以下に示す。カバーテキストとしては本稿のアブストラクトの一部を用いた．置換した表現を [と] で囲んでいる．

ネットワーク基盤の発達等 [によって]，電子的にやり取りされる情報量が目覚ましく増加している．それに伴い，電子化されたコンテンツに対する著作権保護が大きな問題となっている．[そのような] 問題を解決する一つの方法として注目，研究されているのが情報ハイディングである．しかし，[これまで] の情報ハイディングは，画像，音声などに対するものがほとんどであり，テキストを対象にしたものでも，文字間や行間なを微妙に変化させて情報の隠蔽を行うなど，実質的には画像的な扱いをするものがほとんどであった．

図 4: 生成されたステゴテキストの例

4 実証実験

本システムの有効性を実証するため，文書作成に携わる業務をしている被験者複数名に対して，以下の 2 種類の実験を行った．

- カバーテキストと，それから生成されたステゴテキストを比較し，その 2 つの間で日本語として文法的，意味的に等しいかを検討してもらう．(以下，比較実験と呼ぶ)
- ステゴテキストのみを査読させ，その中で使用されている表現が，日本語として正しく，かつ専門文書としての内容に即しているかを検討してもらう．(以下，査読実験と呼ぶ)

以降では各実験の詳細について述べる．

4.1 比較実験

4.1.1 実験概要

比較実験は，生成されたステゴテキストが，もとのカバーテキストと同一の意味を持っているかを検証することを目的とする．文書の種類やサイズの，辞書の内容による影響についても調査するため，カバーテキストとして，5KB 前後のサイズをもつソフトウェアのマニュアル，使用許諾文書をそれぞれ 2 種類ずつと，25KB，50KB のサイズのマニュアルを用意した．文書変換辞書は，一般的な文書に適応するための辞書 (先に定義した一般辞書)，対象となるカバーテキスト中の専門用語も置換の対象とするよう，特定のカバーテキスト用にカスタマイズされた辞書 (以下，専門辞書と呼ぶ)，一般辞書の登録内容を削除し，置換対象を減らした辞書 (以下，縮小辞書と呼ぶ) の 3 種類を用意した．これらを組み合わせ，12 種類のステゴテキストを生成した．

表 1: 実験に使用したステゴテキスト

[1]	[2]	[3]
MS1-t	マニュアル-5KB-1	専門辞書
MS1-c1	マニュアル-5KB-1	一般辞書
MS2-t	マニュアル-5KB-2	専門辞書
MS2-c1	マニュアル-5KB-2	一般辞書
LS1-t	使用許諾-5KB-1	専門辞書
LS1-c1	使用許諾-5KB-1	一般辞書
LS2-t	使用許諾-5KB-2	専門辞書
LS2-c1	使用許諾-5KB-2	一般辞書
MM1-c1	マニュアル-25KB-1	一般辞書
MM2-c1	マニュアル-25KB-2	一般辞書
ML1-c1	マニュアル-50KB-1	一般辞書
ML2-c1	マニュアル-50KB-2	一般辞書
MM1-c2	マニュアル-25KB-1	縮小辞書
MM2-c2	マニュアル-25KB-2	縮小辞書

[1]:ステゴテキスト名 [2]:カバーテキスト名 [3]:使用した辞書
カバーテキスト名は 文書種別 — サイズ (KB) — 識別子の形式で内容を表す．

表 1に示すステゴテキストに対し，各 2 人の被験者にカバーテキスト—ステゴテキストの比較を行ってもらった．制限時間などの制限事項は特に設定しなかった．

4.1.2 実験結果

表 2に比較実験の結果を示す．

これらの結果において，カバーテキストの種類，辞書の登録内容，置換個所数のそれぞれが与える影響について考察する．

- カバーテキストの種類

カバーテキストの種類が，表現の置換においてどのような影響を与えるかを考察する．条件をで

表 2: 比較実験の結果

[1]	[2]	[3]	[4]	[5]	[6]	[7]
MS1-t	50	48	23	1 0	4.3 % 0 %	3 5
MS1-c1	17	17	10	0 0	0 % 0 %	4 5
MS2-t	310	304	123	22 39	17.9 % 31.7 %	5 4
MS2-c1	37	32	16	3 4	18.8 % 25.0 %	5 2
LS1-t	76	50	30	1 0	3.3 % 0 %	5 5
LS1-c1	52	48	26	0 0	0 % 0 %	5 5
LS2-t	154	153	86	4 9	4.7 % 10.5 %	5 4
LS2-c1	39	32	15	0 8	0 % 53.3 %	5 1
MM1-c1	147	144	79	4 2	5.1 % 2.5 %	3 5
MM2-c1	154	152	83	8 15	9.6 % 18.1 %	5 5
ML1-c1	318	312	142	7 4	4.9 % 2.8 %	3 5
ML2-c1	229	224	120	16 19	13.3 % 15.8 %	5 4
MM1-c2	75	73	39	1 0	2.6 % 0.0 %	4 5
MM2-c2	79	72	42	5 8	11.9 % 19.0 %	5 5

[1]: ステゴテキスト名　[2]: 置換可能個所　[3]: 置換個所
[4]: 不一致個所　[5]: 不一致指摘個所　[6]: 不一致指摘率
[7]: 全体評価
(5: 意味はほぼ同じ 3:全体の文意は同じ 1:同じ意味ではない)

きるだけ同じものとするため，辞書に一般辞書を用いて，5KB 程度のサイズのカバーテキストから生成したステゴテキストについて比較する．この条件を満たすステゴテキストは，MS1-c1，MS2-c1，LS1-c1，LS2-c1 の 4 つである．

まず，置換可能である個所は MS1 が 17 個所，MS2 が 37 個所，LS1 が 48 個所，LS2 が 39 個所となっており，全体的に使用許諾文書の方が，一般辞書を用いた場合には埋め込み可能個所が多い．これは，マニュアルはある特定のものについて解説するという性質ものであるため，専門用語が多く出現するということに由来するものと考えられる．実際にそれぞれのテキストにカスタマイズされた辞書を用いた場合には，埋め込み可能個所が飛躍的に増えることからも，このことが言える．

また意味の比較結果についてみてみると，MS1，LS1，はどちらも同一の意味であるという結果が，MS2 については 3〜4 個所の，意味的に不自然な部分が出現していると言う結果となった．LS2 は特殊なケースであり，2 人の査読者のうち 1 人は意味的に同一であるという判断を下しているのに対し，もう 1 人は置き換えた個所の半分以上の部分に対して意味が異なるという判断を下している．全体としての評価も，1 人はほぼ同じ意味であると評価しているのに対し，もう 1 人は同じ意味とは言えないという評価をしている．これは，個人の感性の違いによる部分が大きく，定性的に判断することが難しい．

全体的な評価をすると，一般的な表現の置き換えによって，単純に文章としての意味を保存するという観点からすれば，使用許諾文書の方がマニュアルより，隠蔽情報量，生成されたステゴテキストの質の両方において優れていると言える．

• 辞書の登録内容次に辞書の登録内容という側面から結果についての考察を行う．我々が実験に用いた専門辞書は，一般辞書に専門的な用語を追加登録したものであるため，置き換えの対象とする語が多い．このことが表現の置き換えに与える影響について調べるため，ここでは，カバーテキストが同一で使用する辞書が違うものについて比較することとする．比較するものは，MS1-c1 と MS1-t，MS2-c1 と MS2-t，LS1-c1 と LS1-t，LS2-c1 と LS2-t の 4 組である．

前述のように専門辞書の方が登録単語数が多いため，埋め込み可能個所もそれにしたがった結果となっている．また，これに伴い表現の置換個所も増えたため，意味が不一致である個所が増えている．特に MS2 の対は平均で 8 倍以上，意味が一致していない個所が増加している．しかし，置き換えられた個所に対する意味不一致の指摘個所の割合は LS1 を除いてはあまり変化がなく，全体評価に関しては，平均値が向上している．このことから，辞書の登録内容は，専門用語等，対象となる文書独特のものであっても置き換えるように登録してあったほうがよいということが言える．

• 置換個所数

カバーテキストの文章量が異なる場合と，辞書に登録されている要素数で置換可能個所数は変わってくる．それらを評価するため，同分野の文書で辞書も同一のものを使用するが，カバーテキストのサイズが異なる場合と，これとは逆で，同一なカバーテキストと，登録内容の異なる一般辞書を用いる場合について考察する．

カバーテキストのサイズは異なるが，その分野，辞書が同一なものは，結果中の MS1-c1 と MM1-c1，および ML1-c1 の組，それと，MS2-c1 と MM2-c1，および ML2-c1 の組の 2 つである．前者の組については，特別カバーテキストサイズが大きくなったことの影響はあまりみられず，埋め込み可能個所，不一致指摘個所数等は比例的に推移をした．不一致指摘率，総合評価等の観点からみれば，カバーテキストサイズの小さいものほどステゴテキスト質が良く，サイズが大きくなる

と，ステゴテキストの品質は低下するという傾向が見て取れる．後者の組は，これと逆で，カバーテキストサイズが小さいときほどステゴテキストの質が良くないという結果が出た．

一方，登録内容の異なる辞書を用いて情報の隠蔽を行ったものは，MM1-c1 と MM1-c2，および MM2-c1 と MM2-c2 の 2 組である．これらについてみてみると，必ずしも置換え個所の少ない文書の方がより自然であるとは言えない，という結果が出ている．

これらのことから，置換個所数もステゴテキスト生成の際に何らかの影響があるとは考えられるが，それらは使用する辞書の品質，カバーテキストの品質等によって変わってくるものであると考えられる．

4.2 査読実験

4.2.1 実験概要

査読実験は，ステゴテキスト単体でみた場合，日本語の品質が維持され，かつ専門分野における文書としての正当性を保持しているかどうかを検証するための実験ある．比較実験で用いたカバーテキスト—辞書の組み合わせと同じ組合わせを用い，埋め込みデータだけを変えステゴテキストを生成した．それらのステゴテキストに対し，各文書あたり 2 人づつ文書の正当性を評価してもらった．

4.2.2 実験結果

表 3と 4に査読実験の結果を示す．

本システムにおいて，情報埋め込み，ステゴテキスト生成の仮定において，生成されるステゴテキストの品質に最も大きな影響力をおよぼすものは辞書である．ここでは，査読実験の結果に対して，辞書による表現の置換がおよぼす影響と共に考察する．

まず，比較実験の結果と比べ，全体的に問題指摘個所が増加していることがあげられる．単純に，本実験の問題指摘率と比較実験の不一致指摘率を比較すると，全体の平均値で 14%程，専門辞書を用いいたもの，一般辞書を用いたものはそれぞれ，18%，13%程度づつ値が増加している．この原因として考えられるのは，比較実験において，日本語としては意味が同一であると判定されたものでも，専門的な分野の文書としてみた場合には，意味が異なってくる可能性があるということである．特に使用許諾文書のような法律的な文書に関しては，すでに文書において用いる言葉，言い回し等が固定されていることが多く，こういったものを置換えてしまうと，日本語としての意味は通るが，専門分野の文書としては不自然なものとなってしまう．専門辞書における問題指摘率が高いのは，こういった専門用語も置換えの対象として含まれていることによるものと思われる．

表 3: 査読実験の結果

[1]	[2]	[3]	[4]	[5]	[6]
MS1-t	50	48	26	5 / 12 3 / 3	19.2 % 11.5 %
MS1-c1	17	17	10	4 / 14 0 / 0	40.0 % 0 %
MS2-t	310	304	151	83 / 98 13 / 20	55.0 % 8.6 %
MS2-c1	37	32	17	6 / 12 8 / 10	35.3 % 35.3 %
LS1-t	76	74	39	8 / 10 11 / 14	20.5 % 28.2 %
LS1-c1	52	48	26	1 / 3 2 / 5	3.8 % 7.7 %
LS2-t	154	153	67	9 / 12 28 / 34	13.4 % 41.8 %
LS2-c1	39	32	14	2 / 2 4 / 9	14.3 % 28.6 %
MM1-c1	147	144	71	48 / 52 2 / 6	67.6 % 2.8 %
MM2-c1	154	152	79	2 / 3 35 / 64	2.5 % 44.3 %
ML1-c1	318	312	164	91 / 118 4 / 7	55.5 % 2.4 %
ML2-c1	229	224	117	2 / 2 27 / 153	20.5 % 28.2 %
MM1-c2	75	73	28	18 / 19 0 / 0	64.3 % 0.0 %
MM2-c2	79	72	33	0 / 0 13 / 51	0.0 % 39.4 %

[1]: ステゴテキスト名　[2]: 置換可能個所
[3]: 置換個所　[4]: 不一致個所数
[5]: 問題指摘個所数 (変更分)/問題個所数 (全体) [6]: 問題指摘率

また，表 3を見ると実際に置換えた個所以外にも，問題が指摘されている個所が多いことに気づく．この原因として，置換えにより文脈の解釈の仕方が複数できる可能性があることがあげられる．比較実験の際には，比較するカバーテキストがあったため，あらかじめ文脈はわかっている．しかし，査読実験の場合には，基準となる文書がないため，文脈の解釈の仕方は査読者に委ねられる．複数存在する解釈のうち，カバーテキストがあらわすものと違う意味で解釈をすると，その後方などで意味のつながりが切れ，そこを問題点として指摘されることが起こりえる．マニュアルなどで，置換え個所以外での問題点の指摘が多いのは，マニュアルの方が使用許諾文書に比べ，前に述べたことを踏まえた上で次のことを述べると言った意味的な依存関係が多く存在するためであると考えられる．

表 4に示す査読者側による全体評価は，比較実験に比べ問題指摘率が多い分，全体的に低下し

表 4: 査読実験の結果 (主観的評価)

文書名	専門文書としての評価	日本語としての評価	総合評価
MS1-t	3 - 5	2 - 5	2 - 5
MS1-c1	3 - 5	3 - 5	3 - 5
MS2-t	1 - 3	2 - 4	2 - 3
MS2-c1	3 - 3	3 - 3	4 - 3
LS1-t	4 - 3	5 - 4	4 - 3
LS1-c1	4 - 4	5 - 5	4 - 4
LS2-t	3 - 1	4 - 2	3 - 1
LS2-c1	5 - 4	5 - 4	5 - 4
MM1-c1	2 - 5	2 - 5	2 - 5
MM1-c1	5 - 2	5 - 3	5 - 2
ML1-c1	3 - 5	3 - 5	3 - 5
ML2-c1	3 - 3	3 - 4	3 - 3
MM1-c2	2 - 5	3 - 5	3 - 5
MM2-c2	5 - 2	5 - 2	5 - 2

5：意味はほぼ同じ 3:全体の文意は同じ 1:同じ意味ではない

なお，各欄の a - b は 1 人目が a，2 人目が b という評価をしたことを表す.

ている．特に専門辞書を用いたものについては，比較実験の際には平均で 4.5 という非常に高い値であったのだが，査読実験においては 2.9 とかなり低下している．比較実験の結果に対する考察の際、専門用語も置換えの対象とした方が良いとの結果になったが，これは日本語としての意味の同一性だけを考慮した立場から導き出されたものであり，元のデータの品質を低下させないという情報ハイディングの前提に立つと，専門用語は置換えの対象としない方がよいということになる．

5 まとめ

我々が提案する意味を変えない置換えによる情報ハイディング方式は，評価実験などから，意味を保存し置換えることに成功していると言える．しかし，専門用語などの扱いなどによっては文書の品質が低下することがあるため，この扱いに対する研究が今後の課題となる．

6 謝辞

本研究は情報処理振興事業協会 (IPA) の情報セキュリティ関連事業 (平成 11 年度) 援助により行われました．本研究を進めるにあたって，御助言，御協力頂いた NTT ソフトウェアの若月秀氏，ジャストシステムの宮城裕氏に感謝致します．同じく御協力頂いた三菱総合研究所の川口修司氏，柏木健志氏に感謝致します．並びに，本研究のシステム開発，評価実験に御協力頂いた，松本研究室の池田竜郎氏，赤木健一郎氏に感謝致します．また，本研究における辞書の作成，評価実験に協力して頂いた中川研究室の福田達也氏，保積裕子氏に感謝致します．

参考文献

[1] J. Brassil, S. Low, N. F Maxemchuk, and L. O'Gorman. Hiding infomation documents images. *Conference on Infomation Sciences and Systems (CISS-95)*, March 1995.

[2] J. Brassil and L. O'Gorman. Watermarking document images with bounding box expansion. *Info Hiding 96*, pp. 227–235, 1996.

[3] Mark Chapman and George Davida. Hiding the hidden:a software system for concealing ciphertext as innocuous text. *ICICS'97*, pp. 335–345, 1997.

[4] Peter Wayner. Mimic functions. *Cryptologia*, Vol. XVI, No. 3, pp. 193–214, 1992.

[5] Peter Wayner. Strong theoretical steganography. *Cryptologia*, Vol. XIX, No. 3, pp. 285–299, 1995.

[6] NF Maxemchuk. Electronic document distribution. *AT&T Technical Journal*, Vol. 73, No. 5, pp. 74–80, 1994.

[7] SH Low, NF Maxemchuk, J. Brassil, and L. O'Gorman. Document marking and identification using both line and word shifting. *Infocom 95*, April 1995.

[8] 小俣祐介. テキストへの情報ハイディング方式に関する研究. 横浜国立大学修士論文, 1999.

[9] 松本裕治, 北内啓, 山下達雄, 平野善隆, 今一修, 今村友明. 日本語形態素解析システム『茶筌』version 1.0 使用説明書, 1997.