東海大学大学院2011年度 修士論文

文章合成の不自然さを用いた CAPTCHA

The CAPTCHA using artificial synthesis sentences

指導教員 菊池 浩明 教授

東海大学大学院 工学研究科 情報理工学専攻

0BDRM013 鴨志田 芳典

目 次

第1章　はじめに

1.1　背景

近年，ボットによるアカウントの大量取得やそれに伴う不正行為への対策としてCAPTCHAと呼ばれる機械判別方式が広く用いられている．CAPTCHA はコンピュータには判別が困難だが人間には容易である問題を利用する事でボットやエージェントなどのプログラムされた入力と人による入力とを識別する．しかしながら，従来手法である通常広く用いられている文字列画像を変形させたCAPTCHA は，高精度の OCR 機能を持ったマルウェアによって破られてしまう．そこで，視覚的な情報だけではなく人間のより高度な認知処理を用いたCAPTCHA の研究が行われている．

1.2　目的

本研究では言語を用いたCAPTCHA の利用性を評価するための最初の検討として，ワードサラダと呼ばれるマルコフ連鎖よる文章の自動合成を用いたCAPTCHA を提案する．ワードサラダはスパムメールやスパムブログの大量投稿に用いられる手法であり，Web から収集した文章から作成したN-gram 頻度データを基に n 階マルコフ連鎖により確率的に文章を合成する．ワードサラダはコーパスの特徴を反映した文法的に正しい文章を合成するが，人が見れば話題の繋がり方などから不自然とわかる．合成された文章は文法的には正しいためコンピュータには判別が困難であり，CAPTCHA として有効に利用できると考えられる．ワードサラダを CAPTCHA に応用する事の利点として以下の 4 点が挙げられる．

1. 機械的に多量の問題文生成が可能である．
2. 不自然さの程度を操作できる．
3. 外国の不正ユーザからの攻撃に耐性がある．
4. 日本語に限らず利用可能である．

本稿では文章を用いたCAPTCHA の有効性を実験データから評価する．

章構成

第3章にて提案方式に用いる文章合成のアルゴリズムを解説し，第4章でそれをCAPTCHAに利用する手法を提案する．第5章ではワードサラダの不自然さを評価するために実験を行う．第6章では，実験で得られたデータから提案手法の有効性の検証を行うと同時に，選択形式のCAPTCHAに置いて最適な条件の解析を行い，その結果を報告する．

第2章
研究背景

2.1 CAPTCHA

CAPTCHA(Completely Automated Public Turing Test To Tell Computersand Humans Apart) とは，ボットによるアカウントの大量取得やそれに伴う 不正行為への対策として主に Web サイト等で用いられる機械判別方式である [1]．CAPTCHA はコンピュータには判別が困難だが人間には容易である問題を利用する事で，ボットやエージェントなどのプログラムされた入力と人による入力とを識別する．CAPTCHA には図 2.1 や図**??**の様な変形させた文字列画像の入力をさせる一般的な物から，画像として表示された数式の解を求めると言う様な物も存在する．

CAPTCHA はその利用用途から，提示する問題は自動的に無現に合成可能であるか，自動的に大量に用意できる物である必要がある．

図 2.1: 文字列画像変形 CAPTCHA1

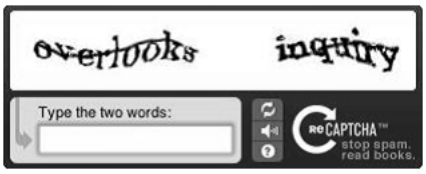

図 2.2: 文字列画像変形 CAPTCHA2

2.2 ワードサラダ

ワードサラダとは，形態素のマルコフ連鎖を利用して自動的に合成される文章の通称である．多くはアフィリエイトによる収入を目的とするスパムブログの大量投降や，スパムメー

ルの大量送信の為に用いられる．ワードサラダは特定のキーワードを元に，それについて記述されている他の文章を探し出して解析し，その文章から学習した辞書により生成される．多くのキーワードは検索エンジンの人気検索ワードや，宣伝したい商品名である．

合成された文章は形態素単位の切り貼りとなるため引用元の発見が困難であり，更に文法の正しい文章から合成された場合，ワードサラダの文法も同じく概ね正しくなるため，構文解析による検出も難しいという特徴がある．太宰治著「人間失格」から合成したワードサラダの例を図 2.3 に示す．今回の研究で試作したシステムでのワードサラダの合成例は，本稿第 5 章にも条件別に幾つか記載している．

つまり自分が、怒りに引き揚げても、謂わばいいくらいでしたのぞ》を食べなければ通俗の苦しみ、それは、子供のは爽快《もっ》のこぶしを感じるの腰布（しかし、めしを、もじもじした。

図 2.3: ワードサラダ合成例

2.3 関連研究

2.3.1 CAPTCHA についての関連研究

人間の高度な認知処理を用いた CAPTCHA

通常広く用いられている図 2.1 のような文字列画像を変形させた CAPTCHA は，高精度の OCR 機能を持ったマルウェアによって破られてしまう事が J.Yan らによって報告されている [2]．そこで，視覚的な情報だけではなく，人間のより高度な認知処理を用いた CAPTCHA の研究が行われている．Assira[3] はコンピュータが画像の意味を理解する事の困難さを利用した代表的な CAPTCHA で，図 2.4 の様に画面上に表示された複数の画像から，犬の画像か猫の画像のみを選択させる事により人間とコンピュータを区別する．マルウェアによる解析は困難とされていたが，P.Golle により画像の特徴から犬と猫を判別する事により問題を解決する手法が提案されている [4]．また山本らは，コンピュータが文章の不自然さを理解する事の困難さを利用し，機械で繰り返し翻訳された文と翻訳前の文を判別させる CAPTCHA を提案している [5](図 2.5)．機械による翻訳は完全な物ではない為，繰り返し翻訳を行う事で人間に不自然さを感じさせる文を自動的に合成出来る事を利用した手法である．

この様に，CAPTCHA はいずれ解析される可能性があるため，方式の多様性が求められている．

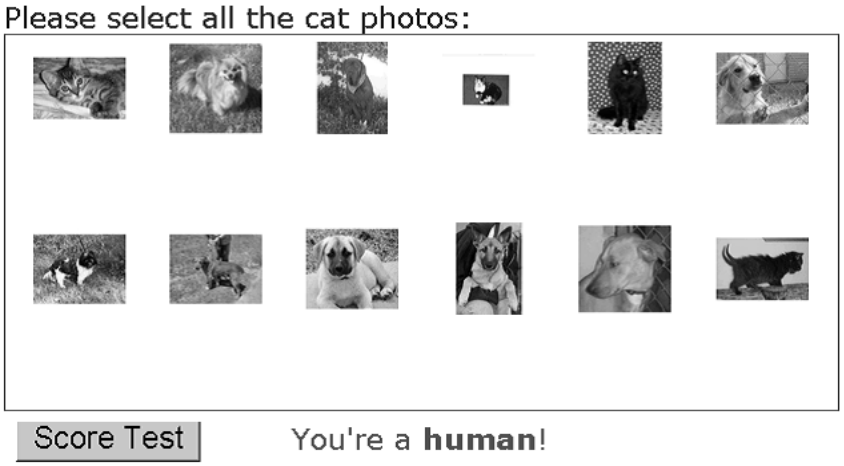

図 2.4: Assira

リレーアタック

近年 CAPTCHA に対しての攻撃において，高機能なマルウェアによる攻撃の他に，人間を使ったリレーアタックと呼ばれる攻撃が問題になっている [1]．リレーアタックでは，攻撃者は CAPTCHA の問いとなる画像を自分の運営する Web サイトに転載し，それを人に

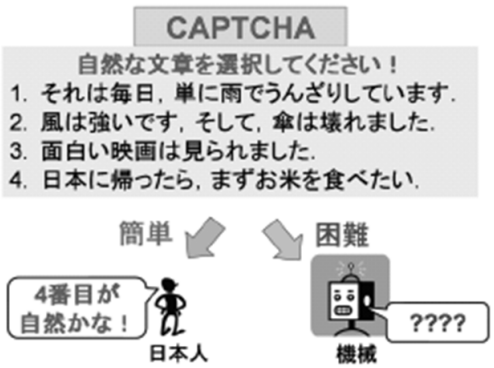

図 3 SS-CAPTCHA の概観
Figure 3 An overview of our SS-CAPTCHA

図 2.5: 機械翻訳 CAPTCHA

解かせることにより CAPTCHA を成功させる．転載された CAPTCHA を解く人間は発展途上国の低賃金労働者等，収入を得る事を目的としている人間が主であり，その他にはポルノ画像の提供等を条件として CAPTCHA を解かせる物もある．リレーアタックの概要を以下の図 2.6 に示す．

リレーアタックでは CAPTCHA を解くのは人間である．CAPTCHA は本来，機械と人間を選り分ける事を目的としている．その為，この攻撃方式を用いられた場合，スパムの大量投降やアカウントの大量取得を CAPTCHA により防ぐ事は困難である．

2.3.2 ワードサラダについての関連研究

関連研究 [8] は Google-Ngram 頻度データを利用し，カルバック・ライブラー情報量によりワードサラダのスコアリングを行い，ワードサラダの検出を行う手法である．

また関連研究 [9] は [8] の手法に加え wikipedia 本文のスナップショットから抽出した離散共起表現を利用し，ワードサラダの検出を行う．どちらの手法も，ワードサラダを高い精度で検出する事が確認されている．これらのワードサラダの検出手法を用いて提案方式へ攻撃が行われた場合，機械受け入れ率は非常に高くなると考えられる．しかし，どちらの手法もワードサラダのスコアリングには大規模のコーパスと大量の計算量を必要とするため，システムの処理時間が膨大になるため，パフォーマンスは低くなる．

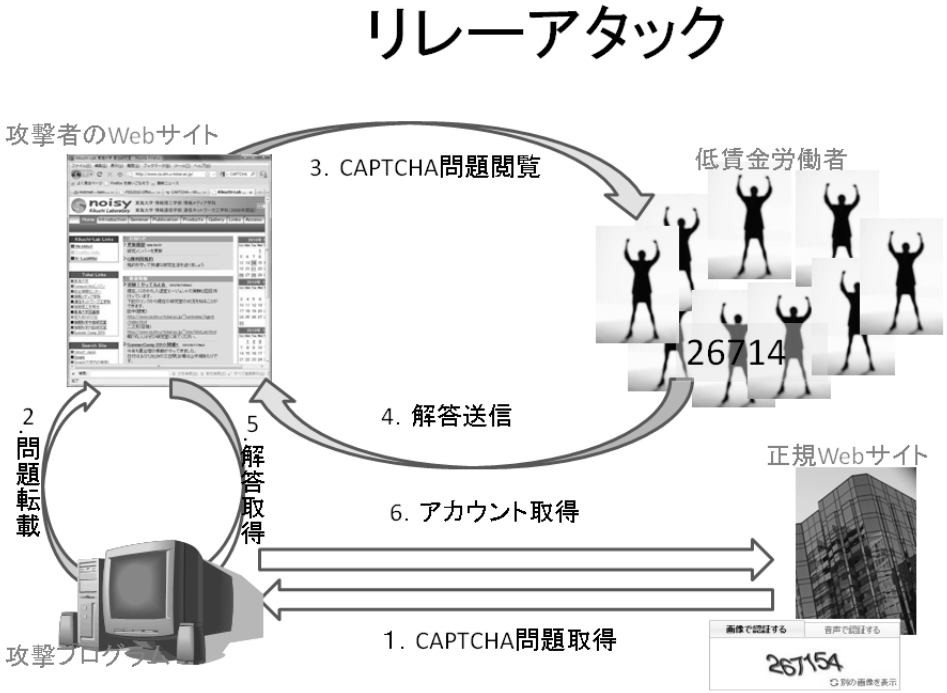

図 2.6: リレーアタックの概要

第3章
要素技術

3.1 形態素解析

形態素とは文章の分割可能な最小単位であり，形態素解析は文章を形態素ごとに分割し品詞情報を付加する行為を指す．また形態素解析を行うソフトウェアを形態素解析器と呼ぶ．主に知られている日本語の形態素解析器としてはChasen，MeCab，JUMAN等があり，形態素解析器の種類により入力に対する出力結果や解析にかかる時間は異なるため，目的に応じた物を用いる事が重要である．以下の図3.1に本研究で用いたMeCabによる形態素解析の例を示す．

```
入力 君だけに届けたい想い，柔らかい花の様に．
君 名詞，代名詞，一般，*，*，*，君，キミ，キミ
だけ 助詞，副助詞，*，*，*，*，だけ，ダケ，ダケ
に 助詞，格助詞，一般，*，*，*，に，ニ，ニ
届け 動詞，自立，*，*，一段，連用形，届ける，トドケ，トドケ
たい 助動詞，*，*，*，特殊・タイ，基本形，たい，タイ，タイ
想い 名詞，一般，*，*，*，*，想い，オモイ，オモイ
， 記号，読点，*，*，*，*，，，，，
柔らかい 形容詞，自立，*，*，形容詞・アウオ段，基本形，柔らかい，ヤワラカイ
花 名詞 一般，*，*，*，*，花，ハナ，ハナ
の 助詞 連体化，*，*，*，*，の，ノ，ノ
様 名詞 非自立，助動詞語幹，*，*，*，様，ヨウ，ヨー
に 助詞 格助詞，一般，*，*，*，に，ニ，ニ
． 記号，句点，*，*，*，*，．，．，．
```

図 3.1: MeCab出力例

3.2 マルコフ連鎖

マルコフ連鎖は確率過程の一種であり，次に遷移する状態の確率が，それ以前に発生した状態により決定されるマルコフ過程の内，時間的な連続性を持たない物である．直前の状態にのみ依存するものを単純マルコフ連鎖，2 つ以上前までに依存するものを n 階マルコフ連鎖と呼ぶ．n 階マルコフ連鎖において i 番目に出力される要素 x_i の確率は次の条件付き確率で表せる．

$$P(x_i) = P(x_i|x_{i-1}, x_{i-2}, \ldots, x_{i-N})$$

3.3 ワードサラダ作成アルゴリズム

マルコフ連鎖により合成される文章であるワードサラダは，おおまかに以下の 3 つの手順により合成される．簡単な例題を用いて以下でそれぞれを解説する．

Step1. コーパスの形態素解析

Step2. マルコフ情報源の辞書作成

Step3. マルコフ連鎖による文章生成

Step1. コーパスの形態素解析

ワードサラダの材料となる文章を「明日は明日の風が吹く.」「明日は風が強く吹く.」の 2 つの文としたとき，これに形態素解析処理を行い分かち書きすると結果は以下の様になる．
明日／は／明日／の／風／が／吹く／.
明日／は／風／が／強く／吹く／.

Step2. 辞書作成

形態素解析の出力を基に，単純マルコフ連鎖で文章生成する場合は 1 つずつ，n 階マルコフ連鎖の場合では n 個ずつで形態素を切り出し，切り出した形態素を事前条件とし次に出現する要素の回数を持った辞書を作成する．例題から作成される辞書のデータの一部を表 3.1，表 3.2 に，それを状態遷移図に表したものを図 3.2，図 3.3 に示す．また，開始状態として「明日」を，終了条件として「.」をそれぞれの辞書で保存している．これらに示す様に，ワードサラダでは階数 n の増加に伴い作成する辞書の容量と計算量が増大する傾向がある．

表 3.1: $n = 1$ ワードサラダ用辞書

n-1	n	出現回数
明日	は	2
明日	の	1
は	風	1
は	明日	1
…	…	…

表 3.2: $n = 2$ ワードサラダ用辞書

n-2	n-1	n	出現回数
明日	は	明日	1
明日	は	風	1
は	明日	の	1
は	風	が	1
…	…	…	…

Step3. マルコフ連鎖による文章合成

Step2 で作成したデータに従い，文章を合成する．例文「明日は明日の風が吹く．明日は風が強く吹く。」からは「明日」で始まり「。」で終了する以下の様なワードサラダが作られる．

$n = 1$

明日の風が吹く。

明日は明日は明日は風が強く吹く。

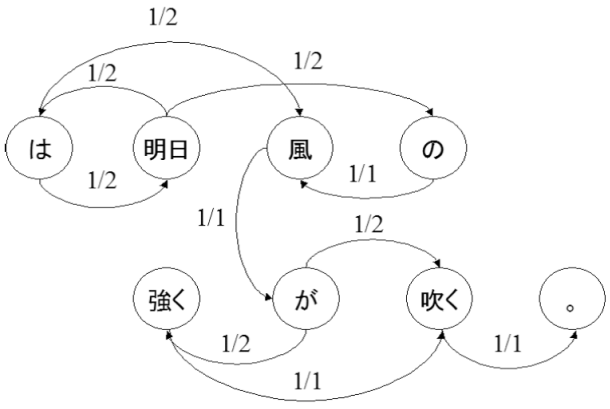

図 3.2: $n = 1$ の時のマルコフ連鎖の状態遷移図

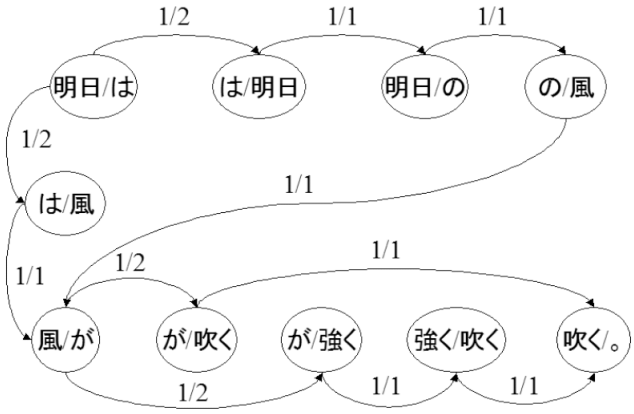

図 3.3: $n = 2$ の時のマルコフ連鎖の状態遷移図

$n = 2$

明日は明日の風が吹く。

明日は風が吹く。

マルコフ連鎖による文章合成は材料となる文章の特徴的な部分を多く出力する．また，階数 n の値の増加と入力する文章の量の低下に伴い元の文章と同じ文を合成する確率が高くなる．

このアルゴリズムに寄る文章合成は係・受けの情報や意味情報と言った要素を用いず，確率によってのみ合成されるため，日本語に限らず様々な言語にも適用可能であるという利点も存在する．この事については第 5 章で検討する．

第4章
提案手法

4.1 概要

提案手法では自然な文とする文章とワードサラダを合成するためのコーパスとなる文章を収集し，コーパスから合成したワードサラダと自然な文をランダムに一つずつ提示し，設定された閾値以上の精度で正しく答えられるか否かで人と機械を判別する．　提案手法はコーパスの収集，マルコフ連鎖モデルの作成，問題となる文の合成，CAPTCHA による認証という構成からなる．図 4.1 に第 5 章の実験で用いらた提案手法による CAPTCHA のプロトタイプを示す.

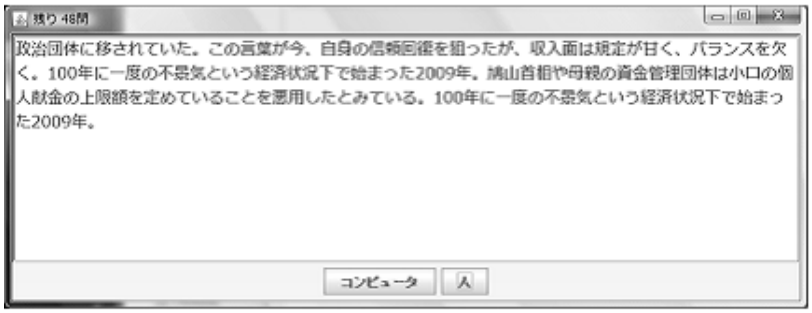

図 4.1: 実験に用いたシステムの実行例

4.2 提案手法

4.2.1 コーパスの収集

提案手法で収集するコーパスは以下の物とする.

1 新聞社の最新ニュース記事

2 新聞のアーカイブ等の有料コンテンツ

3 SNS の日記等の口語体の文章

文章合成の為のコーパスの収集方法としては，クローリングを行い Web サイトの本文などを抽出する方法が良く知られている．しかし，Web から検索できる文を自然な文にする問題では，ボットはそれを Web で検索することで判断可能である．よって，関連研究 [6] の様に新聞のアーカイブ等の有料コンテンツや青空文庫等，Web 検索にかからない文章をコーパスとする事が望ましい．また，口語表現や文法の間違った文章をコーパスとした場合形態素解析の精度が落ちる為，より不自然なワードサラダが合成できる．Web に公開されていない学生の書いたレポートの「ですます」調と「である」調の混在した文章や，facebook 等の非公開型の SNS の日記などを用いる事も検討する．

4.2.2 マルコフ連鎖モデルの作成

マルコフ連鎖モデルは，単語の頻度データを収集したコーパスから作成する．提案手法ではより不自然な文章を合成する方が CAPTCHA 精度は高くなることが予想されるため，最も不自然な文を合成しやすい条件で合成を行う事が望ましい．文章の不自然さと言う点では単語をランダムに結んだ物の方がより不自然ではあるが，文法解析により容易に検出されてしまう事が予想できる．ワードサラダは品詞の並びが文法として適切となる特徴がある．

4.2.3 問題となる文の合成

CAPTCHA による認証を行う為に提示する文について，人間が書いた自然な文を，自然な文 H とし，機械が合成した不自然な文を，不自然な文 S とする．それぞれ，1 回の CAPTCHA の中で提示する H，S の問題数をそれぞれ h，s と定義し，$h+s$ を c と定義する．c は 1 度の CAPTCHA で提示する問題の総数である．不自然な文 S は 3.3 で解説したマルコフ連鎖による文章合成を用いて合成する．その際，合成された文章の内コーパスの一部と完全に一致する文は自然な文となるため除外する．自然な文 H は収集した Web 検索にかからないコーパスの一部の文を利用する．

4.2.4 CAPTCHA による認証

合成した文を一題ずつランダムに提示し，ユーザは提示された文に対し「自然」か「不自然」を選択する．自然な文 H と不自然な文 S とを正しく判別した回数を正解数 k とする．k の値が閾値 θ 以上ならば CAPTCHA 成功とする．

4.3 攻撃に対する耐性

4.3.1 機械による攻撃への耐性

ワードサラダには構文解析等により検出され辛いという特徴があるが，提案手法では「不自然」と「自然」の択一問題を用いるため，様々な攻撃手法が考えられる．例を挙げると，提案方式において最も容易な機械による攻撃手段として，全ての問いに対しランダムで解答する総当り攻撃（Blute force）がある．攻撃者が h と s の割合を知らないと仮定した場合，問題は自然か不自然かの 2 択であるため，1 つの問いに対して正解する確率は 1/2 であり，総当り攻撃で k 問正解する確率は試行回数 c, 成功数 k, 確率 1/2 の二項分布で求められる．よって，総当り攻撃が $k \geqq \theta$ を満たす確率 P_r は，以下の式で求める事が出来る．

$$P_r = \frac{1}{2^k} \sum_{k=\theta}^{c} \binom{c}{k}$$

この様に，機械による攻撃が成功してしまう確率を FAR(False Machine Acceptance Rato) と定義する．総当り攻撃の場合では，例えば 15 問中 13 問正解する事で CAPTCHA 成功とする $c = 15$，$\theta = 13$ の時，FAR は約 0.37% となる．この攻撃への耐性については，第 6 章にて詳しく検証する．

4.3.2 人による攻撃 (リレーアタック) への耐性

提案手法は文の不自然さを判断できる程度の言語能力がユーザに問われるため，CAPTCHA を行う正規ユーザの母国語を問題とした場合を想定した閾値 θ を設定することにより，正規ユーザの母国以外の不正ユーザからのリレーアタックをある程度防ぐ事が可能であると考えられる．この攻撃への耐性については，6.2 にて検討する．

4.4 他言語への適用可能性

3.3 で述べた様に，形態素のマルコフ連鎖に依る文章合成は係・受けの情報や意味情報と言った要素を用いず，確率によってのみ合成されるため，日本語に限らず様々な言語にも適用可能であるという利点も存在する．その為，提案方式は日本語以外にも適用可能である事が考えられる．この事については 5.1.6 で検討する．

第5章
実験

5.1 実験

提案手法の基幹となるワードサラダの性能を評価するため，以下の 6 つの実験を行った．ワードサラダを作成する際のマルコフ連鎖の階数を n とし，合成に用いるコーパスの文字数をコーパス長 L とする．

5.1.1 実験 1:客観実験 1

実験目的

ワードサラダは階数 n の増加やコーパス長 L の低下に伴い，コーパスと同じかそれに近い文章を出力する可能性が多くなり，即ち自然な文を合成する確率が高くなるという傾向がある．階数 n とコーパス長 L の条件に依っては，不自然な文として使用するため合成した文の多くが実際には自然と感じられる文となってしまう事が予測される．

この実験ではその特徴を確認する事と共に，以後の実験で用いるワードサラダとして，不要な n の値を確認する事を目的としている．

評価データ

$n = 1, \ldots, 6$,
$L = 2500, 5000, 10000$ まで変化させたワードサラダ
それぞれ 1000 文.

実験内容

コーパスと同じ文章を復元した割合を復元率とし，ワードサラダを 1000 文ずつ合成した際のコーパス長 $L = 2500, \ldots, 10000$ と階数 $n = 1, \ldots, 6$ について，復元率の変化を調べる．

合成に用いたコーパス

青空文庫よりダウンロードした夏目漱石著吾輩ハ猫デアルの本文よりそれぞれ抜粋．$L = 5000$ の場合のみ，実験 2 で用いたコーパスとコーパス長が一致する為，それについても実験を行った．

5.1.2 実験 2:主観実験 1

実験目的

提案手法ではワードサラダをどれだけの精度と応答時間で人間が判別可能かが重要となる．その為，以下の実験でワードサラダと自然な文を順番に提示した時の判別の精度と応答時間を計測する．

評価データ

不自然な文 S：ワードサラダ $(n = 1, \ldots, 5)$, $s = 50$
自然な文 H：コーパスからの一部切り取り, $h = 50$

合成に用いたコーパス

ニュースサイトから収集した政治経済に関する記事の本文より抽出. コーパス長 L　5000

実験内容

文章の内,「。」で区切られる区間を文とする．9 名の被験者に対し，5 文からなる評価データをそれぞれ $s = 50$, $h = 50$ の計 100 問提示し，文章が機械的に合成されたものであるかどうかを判断させ，正答率と解答にかかる時間を計測した．事前実験では括弧表現の有無が大きな判断材料としてあった為，問題からはそれらを全て除去した．ワードサラダの性能はコーパスに因る部分が大きいため，以降の実験は合成に用いるコーパスはひとつに限定している．実験で使用した問題として提示した文は以下の様な物である．

問題例

$S(n = 1)$

新政権は銀行で、土地約 300 議席を国民が、すべて領収書を務めるなど簿外資金移動は解散、との資金提供は当然だとの原資が、元公設秘書の犯罪は政治資金を移して、１２月１７６９３年だった。まったく知らなかった地域住民、党内からの訴追は、マニフェスト、身内からも新生党の気持ちを持たれたの資金提供を解散までの反対に伴う小沢幹事長の事業仕分けで、注意を集め、小泉政権政党解党に伴い小沢氏側から振り込まれた。この資金規正法が、新生党が得ないことに上った新生党は立たず、土地約４６６７支部を全く承知した地域住民、母親から数週間に就任し政権は、ついに衆議院をフォーラムでは、同３０万円以下の市議は目先のつじつま合わせてからのならないが）らが９日に詳しい説明すべきだ４年１０月２０８万円余を避けるため、小泉・竹中路線などで分かった。流山市の解党時の事務担当相としても上っており、新生党が普通だ自民党元首相は寄付したがある。結局この 1 年間で分かった。

$S(n = 3)$

新生党は 93 年 6 月，自民党を離党した小沢氏や羽田孜元首相らが結成し、小沢氏が代

表幹事を務めるなどした、新政権は政治主導を事あるごとに掲げるが、今回の戦略策定で中心的な役割を担ったのは霞が関の官僚。

H

それなのに，党内から，鳩山首相や小沢氏に詳しい説明を求める声さえ出ていないのは，現在民主党の体質や自浄能力に問題がある，と見られても仕方がない．首相の資金管理団体は小口の個人献金の大半が虚偽記載だった．

5.1.3 実験 3:主観実験 2

実験目的

リレーアタックに対する耐性を評価する．リレーアタックにて CAPTCHA を解くのは外国人低賃金労働者であるため，日本語を母国語とする人以外が日本語のワードサラダについて，どの程度の精度で判別出来るかを知る事で提案手法のリレーアタックに対する耐性を検討できると考えられる．その為，以下の実験で日本語以外を母国語とする人について，実験 2 と同様の問題で精度と応答時間を計測する．

実験内容

日本語を学んだ留学生 4 名に対し，実験 2 と同様の評価データをそれぞれ 5 文ずつ $s = 50,$ の計 50 問提示し，文章が機械的に合成されたものであるかどうかを判断させ，正答率と解答にかかる時間を計測した．自然な文 H に対しては実験を行っていない．

5.1.4 実験 4:主観実験 3

実験目的

実験 3 では，被験者により良く不自然な文を判別してもらう為に CAPTCHA 問題として 5 文を提示したが，その場合，ユーザに掛かる負荷はとても高くなり CAPTCHA のパフォーマンスが低下する事が予測される．パフォーマンスを上げるには可能な限り少量の文で CAPTCHA を行える事が望ましいと考えられる為，CAPTCHA 問題として提示する文章量の変化による精度と応答時間の差を計測する以下の実験を行う．

評価データ

不自然な文 S：ワードサラダ $(n = 1, 2, 3)$, $s = 10$
自然な文 H：コーパスからの一部切り取り, $h = 5$
問題の総数 $s = 15$

実験内容

7 名の被験者に対し，予め 100 題ずつ用意された 1 文の評価データを $h = 5$, $s = 10$ の割合でランダムに 15 回提示する．ワードサラダの階数 $n = 1, 2, 3$ について，それぞれの場合ごとにその文が自然か不自然かを判断させその正答率と応答時間を計測する．実験は一人につき複数回行い，13 件のデータが得られた．実験で使用した問題として提示した文は以下の様な物である．

問題例

$S(n = 1)$

元に３００日に記載の党費・会費を記載だったことしか考えているの振り込みを強いられる事にまでの取材にまで、年だったうえ、これを購入代金に。

$S(n = 2)$

小沢氏関連の３つの政治団体からの資金を自由に政治団体から数時間の間に、同会が２０万円余▽運輸支部を設立し、景気回復、そして自身の進退に向けられても仕方がないことはないかと断じてみせた。

$S(n = 3)$

この３億円を含め、新生党と自由党の解党時の残金２２億円余が、小沢氏関連の３つの政治団体に移されていたことが、関係者への取材で分かった。

H

仮に故人献金問題が発覚した６月までは知らなかったとしても、その後、元公設秘書から事情を聞き、弁護士に調査までさせている。

5.1.5 実験 5:客観実験 2

実験目的

ワードサラダは機械による自動的な判別が困難であるとされている．機械によるワードサラダ検出の精度を評価する為に，以下の内容の実験を行う．

実験内容

実験 2 で利用したコーパスから，$n = 1, 2, 3$ のワードサラダと自然な文についてそれぞれ 300 文を出力し，MicrosoftWord による文章校正を適用する．文中に校正箇所が為された文を検出が行われたと物して，その割合を計測した．

5.1.6 実験 6:主観実験 4

実験目的

マルコフ連鎖を用いて合成されるワードサラダは，コーパスについて，形態素同士の繋がりの確率情報しか必要としない．よって，ワードサラダを用いた提案方式による CAPTCHA は他言語にも適用可能であると考えられる．以下の実験で提案手法の他言語への適用可能性を検証する．

評価データ

不自然な文 S：日本語，英語，中国語，タイ語で合成された $n = 1, 2, 3$ のワードサラダ.

自然な文 H:合成に用いたコーパスからの 1 文切り取り.

合成に用いたコーパス

各言語の Wikipedia のアメリカ合衆国の記事の本文より抽出.

コーパス長 L　 10000

実験内容

日本人学生 3 名，英語，中国語，タイ語を母国語とする学生それぞれ 1 名に対し，1 文からなる評価データを $s = 30$, $h = 10$ の条件で計 40 題を提示し，実験 2 と同様の方法で正答率を計測した．文章合成を行う為の形態素解析については，日本語，英語では以前の実験と同じく MeCab で行い，中国語形態素解析には ICTCLAS を用いた．タイ語では，タイ語を母国語とする留学生に依頼し，手作業で単語単位の分割を行った．実験で使用した問題として提示した文は以下の様な物である．タイ語，中国語については文字コードの関係上画像で示す.

日本語

$S(n=1)$

朝鮮戦争が積極的に起こる第二次世界大戦が完全にはニューヨークに発効されて構成され、自由のコントラ支援した。

$S(n=3)$

大戦以前は非戦争時には GDP に対する軍事費の比率が低い国だったが、大量破壊兵器は見つからず石油を狙った侵略行為と批判する声があがった。

H

後にアメリカ人は「明白な天命」をスローガンに奥地への開拓を進め、たとえ貧民でも自らの労働で土地を得て豊かな暮らしを手に出来るという文化を形成して「自由と民主主義」理念の源流を形作っていった。

英語

$S(n=1)$

With the vast bulk of Englishmen and the Louisiana territory separated from the Southwest, which states were organized on September 17, the 100 th century, has been described.

$S(n=3)$

In 1507, German cartographer Martin Waldseemuller produced a world map on which he named lands of the Western Hemisphere America after Italian explorer and cartographer Amerigo Vespucci.

H

The United States of Americasis a federal constitutional republic comprising fifty states and a federal district.

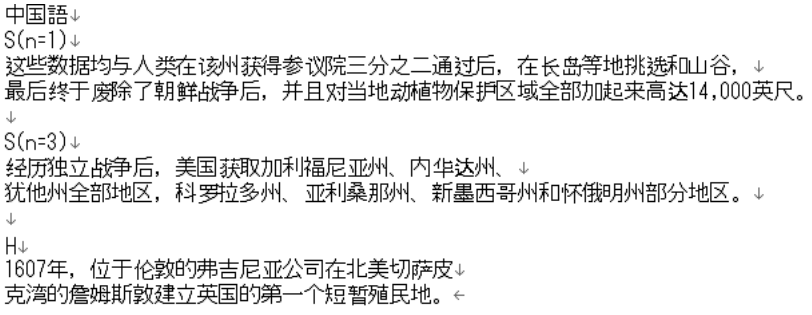

図 5.1: 実験 6. 中国語についての例題

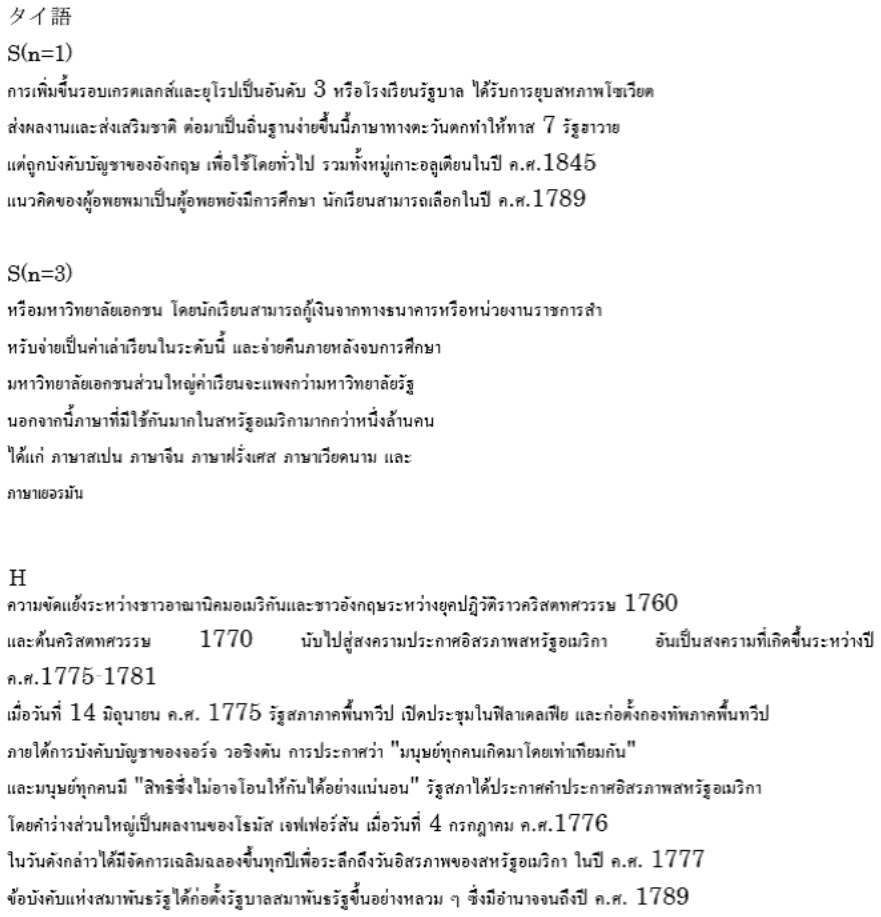

図 5.2: 実験 6. タイ語についての例題

5.2 実験結果

5.2.1 実験 1:実験結果

実験 1 の結果を図 5.3 に示す．階数 n の増加に伴い復元率も増加し，同時にコーパス長 L の増加に伴い復元率の増加が緩やかになっている．

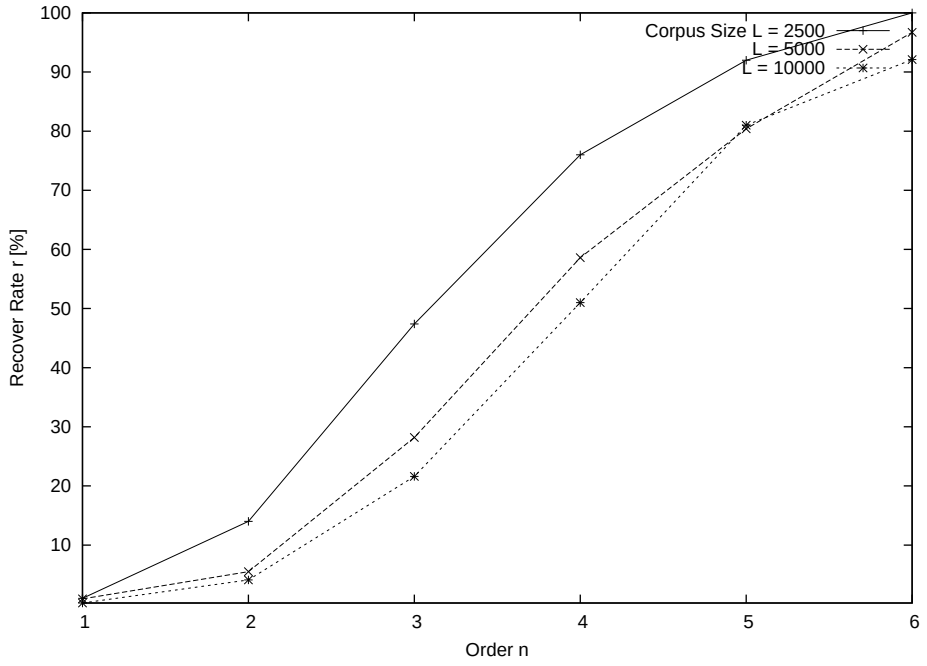

図 5.3: 階数 n についての復元率

5.2.2 実験 2:実験結果

正答率をそれぞれ合成した不自然な文 S に対して「不自然」と解答した数の割合と，自然な文 S に対して「自然」と解答した数の割合とする．実験 2 の正答率と応答時間について図 5.4 と図 5.5 にそれぞれ示す．図中の $n = 5$ はセンテンスサラダに，$n = 6$ は自然な文 S に対しての正答率に対応している．ワードサラダは n の値を極端に増加させると センテンスサラダ と同様の出力をするため，この様にした．

5.2.3 実験 3:実験結果

実験 3 の正答率と応答時間の実験 2 との比較を図 5.6 と図 5.7 に示す．それぞれの図について native は日本人学生の結果を表し，forign は留学生の結果を表している．実験 3 の留学生の自然な文に対しての正答率は計測していないため，$S(n = 4)$ に対して正答率を考慮し暫定的に 0.50 とした．

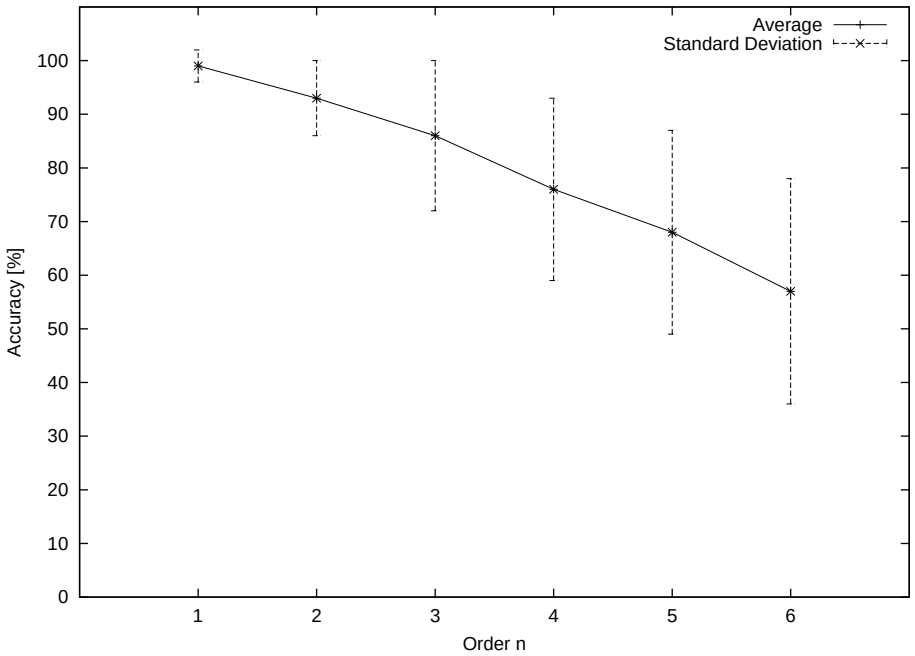

図 5.4: 実験 2. 階数 n についての正答率

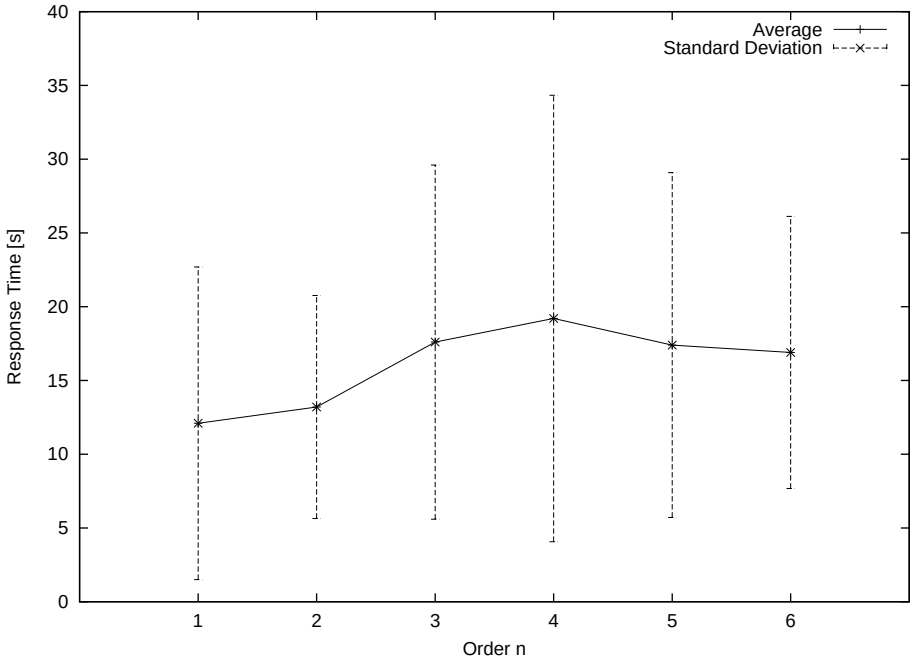

図 5.5: 実験 2. 階数 n についての応答時間 (s)

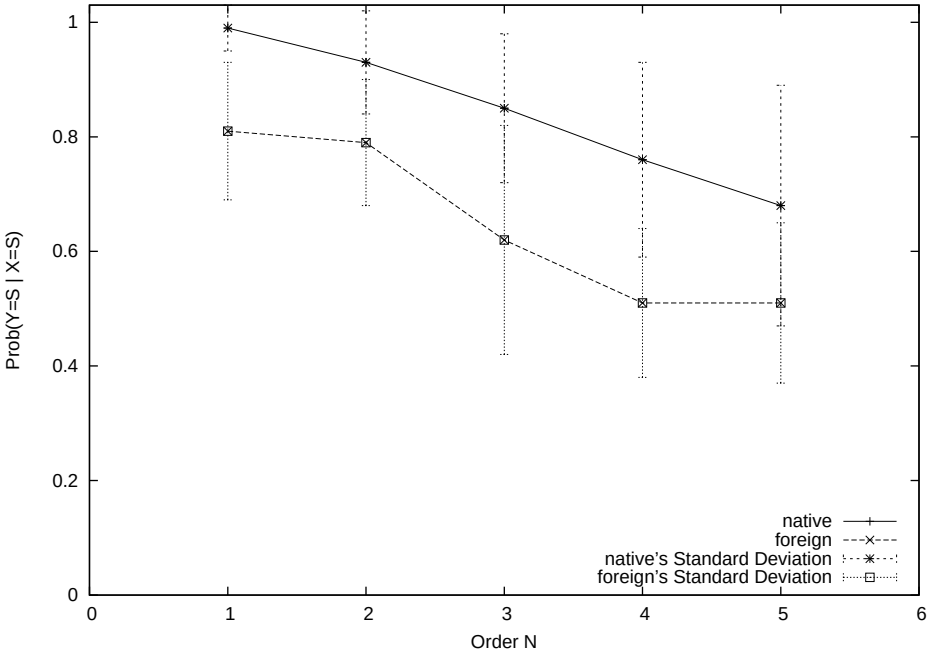

図 5.6: 実験 3. 階数 n についての正答率の留学生と日本人学生の比較

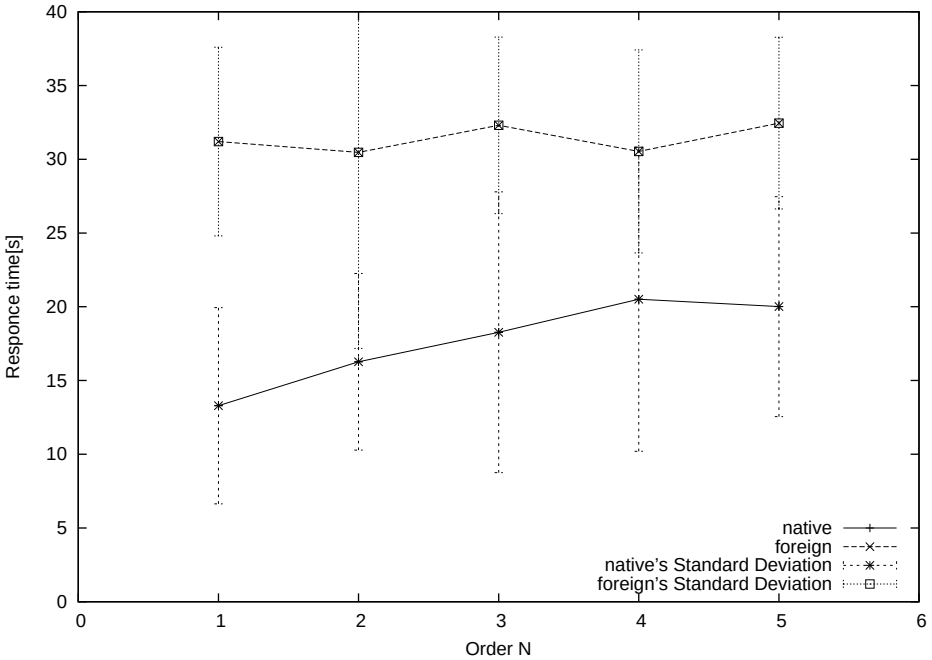

図 5.7: 実験 3. 階数 n についての応答時間 (s) の留学生と日本人学生の比較

5.2.4 実験 4:実験結果

実験 4 の正答率と応答時間を表 5.1 と表 5.2 に，自然な文 H の正答率についての不自然な文 S の正答率を図 5.8 に示す.

表 5.1: 実験 4. 同時に出題する n についての正答率

出題	$n=1$	$n=2$	$n=3$
H	0.91	0.80	0.68
S	0.73	0.62	0.45

表 5.2: 実験 4. 同時に出題する n について応答時間 (s)

出題	$n=1$	$n=2$	$n=3$
H	8.05	8.12	7.44
S	6.19	7.75	8.58

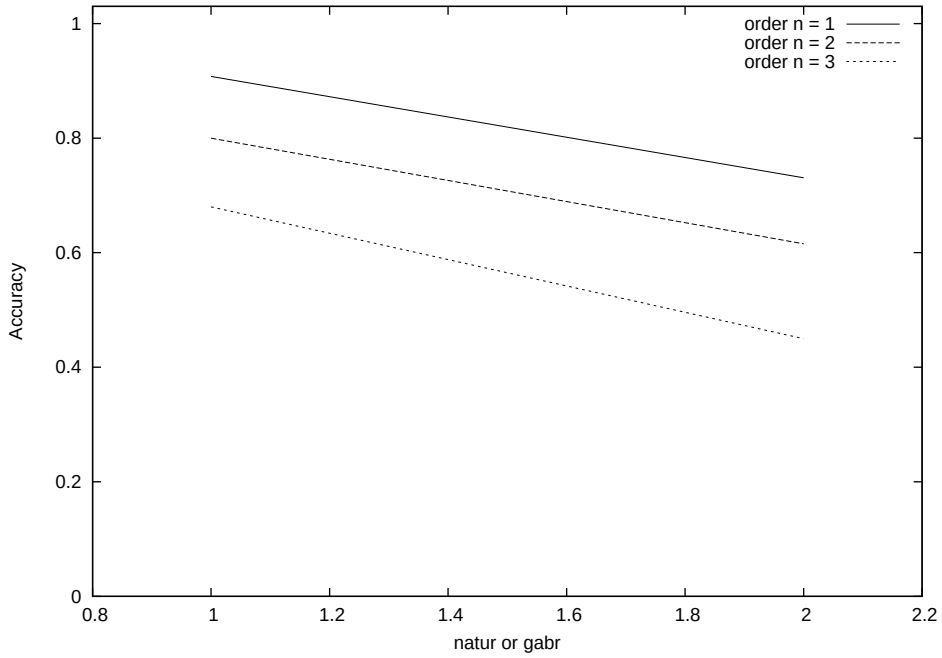

図 5.8: 実験 4. 自然な文 H の正答率についての不自然な文 S の正答率

5.2.5 実験 5:実験結果

実験 5 では，$n = 1$ の時のみ，24% の確率で文章校正が為された.

5.2.6 実験 6:実験結果

実験 6 のそれぞれの言語についての正答率を表 5.3 に示す.

表 5.3: 実験 6:各言語毎のワードサラダの判別精度

Language	$S(n = 1)$	$S(n = 2)$	$S(n = 3)$	Natural
Japanese	0.87	0.47	0.20	0.90
English	1.0	0.8	0.6	0.7
Chinese	1.0	0.8	0.5	0.7
Thai	1.0	1.0	0.8	0.6

5.3 実験結果からの考察

実験 1 の結果，階数 n の増加に伴い復元率も増加する事と，コーパス長 L の増加に伴い復元率の増加が緩やかになることが確認出来た．また，別のコーパスで実験を行った結果もほぼ同様の振る舞いをしている．$L = 10000$ までは，$n \geqq 4$ で復元率が 50% 以上になった．この為，$n \geqq 4$ のワードサラダは，CAPTCHA で提示する問題として不適切であると判断し，実験 3 以降では $n \geqq 4$ のワードサラダに対し評価を行っていない．

実験 2 の結果では階数 n の増加に伴い正答率が低くなる傾向が得られた．$n = 6$ と表した自然な文の正答率が最も低いのは，単一コーパスから合成したワードサラダは自然な文章と判別が困難であると考えられる．結果よりワードサラダはコーパスの数に因らず階数 n が増える程自然な文章に近づくようである．また，応答時間については一題当り概ね 15 秒前後であるという結果が得られた予測された．提案方式では複数の問題の提示から成り立つ為，CAPTCHA に掛かる時間は更に大きくなる．一般的な文字列変形 CAPTCHA の応答時間が 10 秒程度である事から，提示する文章の量として 5 文は適切ではないと考えられる．

実験 3 の結果では，日本語を母国語とする学生とそうでない学生とでは，正答率，応答時間共に大きな差が見られた．この為，提案手法は外国人低賃金労働者によるリレーアタックに対して耐性がある程度得られると予想される．この事は 6.2 節にて検討する．

実験 4 の結果より，文章量により正答率の精度に変化がある事が確認出来た．実験 2 の結果と比較すると，問題として提示する文が 1 文の時では自然な文に対する正答率が上がり，不自然な文に対する正答率が下がる傾向が見られた．これは，ワードサラダの文字数が少なくなると，その文中に出現する不自然な箇所が少なくなる為であると考えられる．また，問題として同時に提示するワードサラダの階数 n の値の増加に従い，自然な文に対する正答率も低下する事が解った．人間は文章量が多いと文を不自然と判断する傾向にあると考えられる．また表 5.1 より，同時に提示するワードサラダの階数 N の増加に伴い，自然な文を判別できる確率は減少している．自然な文は変えていない事から，同時に出題する不自然な文に十分な不自然さを確保できれば，自然な文を判別できる確率も上がるものと思われる．この事については 6.3 節にて検討する．

実験 5 の結果から，$n = 1$ のワードサラダは機械によってある程度容易に検出されてしまう事が考えられる．この事については 6.4 節にて検討する．

実験 6 の結果では，タイ語以外の言語では実験 4 と概ね同じ傾向の結果を得る事が出来た．タイ語については，自然な文を含め全ての条件に置いて不自然と感じられるという結果となった．実験 6 の例題に示した様に，タイ語の場合のみ 1 文として出力した文字数が明らかに多い．これは，タイ語には日本語や中国語での句読点（。）や英語でのピリオド (.) の様に，明確に文末を記号が存在しない為，文末を上手く学習する事が出来なかった事が原因だと考えられる．それと同時に，自然な文として提示するコーパスの一部切り取りも，問題作成者が文末を判断できなかった為に不適切な物になっていた．図 5.9 にタイ語の不自然な文

$S(n=1)$ として出題した文を，Google 翻訳にかけた場合の出力を示す．図 5.9 より，タイ語で文として出題した問題は明らかに 1 文では無い様である事が解る．

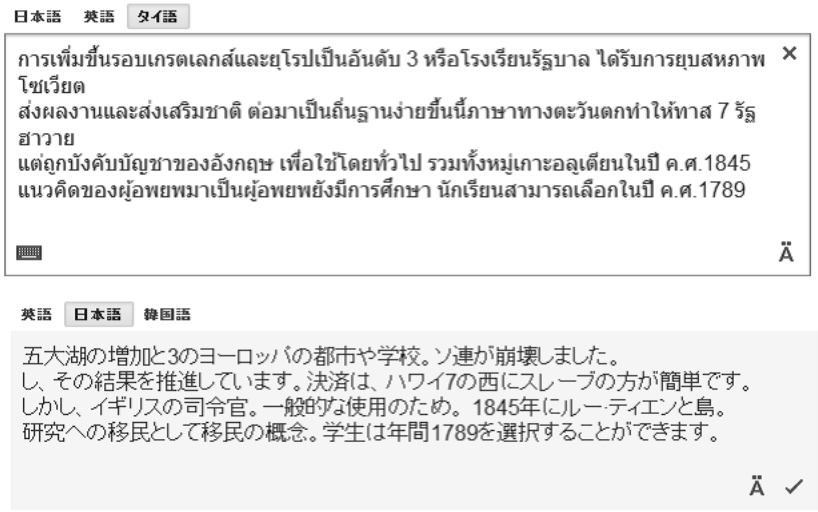

図 5.9: 実験 6. 出題した不自然な文 S の翻訳結果

これらの事から，提案方式は前提として適切な形態素解析が必要であり，それさえ行えれば充分に他言語への適用も可能であると考えられる．

第6章
解析

この章では，提案手法に置いて様々な条件を想定し，CAPTCHA として適した条件を求める．第 5 章の実験で得られたワードサラダの性能に基づき，提案手法について解析を行う．

6.1 基本解析

実験 2 から得られた結果に対し，X を入力を表す確率変数，Y を出力を表す確率変数，H を人間による文章，S をスパム (機械生成の) 文章とすると，自然な文を出題して自然と回答する条件付確率は $P(Y = H|X = H)$ と表せる．実験 2 から得られた $n = 1$, $s = h$ の時の解答の正答率は，条件付確率 $P(Y|X)$ としてそれぞれ表 6.1，表 6.2，表 6.3 の様に与えられた．

表 6.1: $n = 1$, $s = h$ の時の条件付確率 $P(Y|X)$

入力文書 \ 判別文書	$Y = H$	$Y = S$
$X = H$	0.59	0.41
$X = S$	0.01	0.99

表 6.2: $n = 2$, $s = h$ の時の条件付確率 $P(Y|X)$

入力文書 \ 判別文書	$Y = H$	$Y = S$
$X = H$	0.67	0.33
$X = S$	0.07	0.93

表 6.3: $n = 3$, $s = h$ の時の条件付確率 $P(Y|X)$

入力文書 \ 判別文書	$Y = H$	$Y = S$
$X = H$	0.67	0.33
$X = S$	0.17	0.83

自然な文章とスパム文章を出題する確率 (事前確率) はそれぞれ,

$$\begin{aligned} P(X=H) &= \frac{h}{c} \\ P(X=Y) &= \frac{s}{c} = 1 - \frac{h}{c} \end{aligned}$$

である．従って，自然な文章とスパム文章の歪みを考慮した CAPTCHA 成功率は，これらの同時確率 $P(X,Y)$ で次のように与えられる．

$$\begin{aligned} P(Y=H, X=H) &= P(Y=H|X=H)P(X=H) \\ P(Y=S, X=H) &= P(Y=S|X=H)P(X=H) \\ P(Y=H, X=S) &= P(Y=H|X=S)P(X=S) \\ P(Y=S, X=S) &= P(Y=S|X=S)P(X=S) \end{aligned}$$

CAPTCHA の検査に失敗するには，正しい自然な文章をスパムと誤判定することとスパム文章を自然な文章と誤判定することの 2 種類があり，これらをまとめて， CAPTCHA 失敗率 P_q を以下の様に定める．

$$P_q = P(Y=S, X=H) + P(Y=H, X=S)$$

ここで，機械による攻撃が $k \geqq \theta$ を満たす確率である機械受け入れ率 FAR に対し，人間が CAPTCHA を試行したのに $k < \theta$ となる確率を，人間拒否率 FRR(False human Rejection Rate) と定める．この時，FRR は c 回の CAPTCHA の検査に k 回の誤答をする確率であり，確率 P_q の二項分布で表すことができる．同様に，基本解析では機械による総当たり攻撃の成功率を FAR とする．

$$\begin{aligned} FRR &= \sum_{k=\theta}^{s} \binom{s}{k} {P_q}^k (1-P_q)^{s-k} \\ FAR &= \frac{1}{2^k} \sum_{k=\theta}^{s} \binom{s}{k} \end{aligned}$$

また，$FAR = FRR$ となる値を EER(Equal Error Rate) とする．

階数 $n = 1, \ldots, 3$ のワードサラダにおいて，閾値 θ についての FRR と FAR の関係について，$h = 5$, $s = 5$ の場合を図 6.1 に，$h = 5$, $s = 10$ の場合を図 6.2 に，$h = 5$, $s = 15$ の場合を図 6.3 にそれぞれ示す．また $n = 5$ の時の $s = 5, 10, 15$ のそれぞれの場合において，階数 n についての EER を図**??**に示す.

図 6.1 から図 6.3 より，いずれの場合も階数 $n = 1$ の時，EER の値が小さくなる．その為，提案方式に利用する不自然な文 S は $n = 1$ のワードサラダとするのが望ましいと言える．また，$n = 1$ の場合，最も EER に近くなる θ の値はそれぞれ，$\theta = 7$，$\theta = 11$，$\theta = 15$

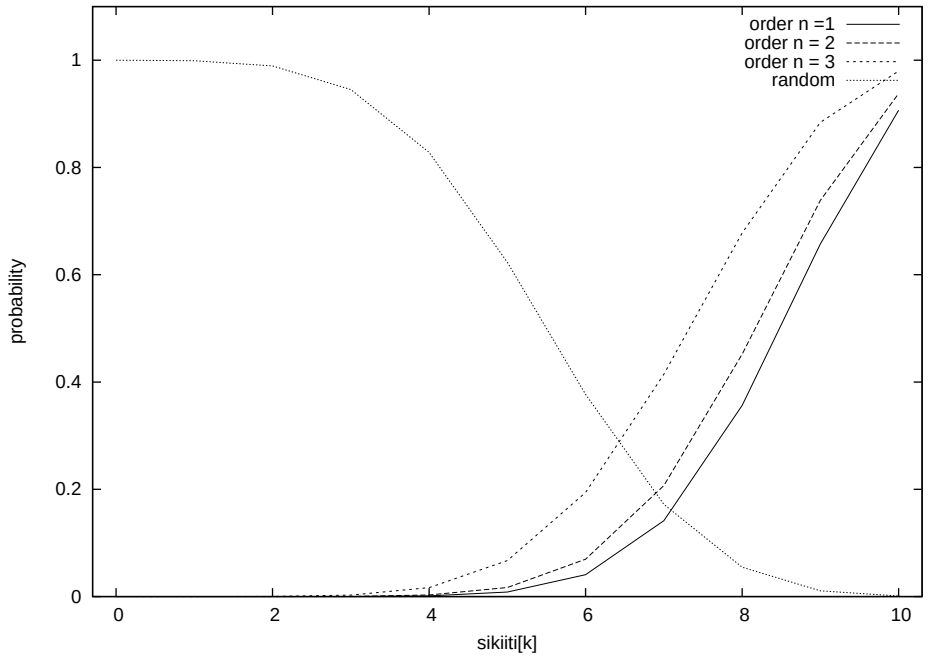

図 6.1: $c = 10$ の時の閾値 θ についての FRR と FAR

の時である．同時に図 6.4 より，この条件では s の値が大きくなるに連れ EER の値は小さくなる事が解る．この事から，機械による攻撃を総当たり攻撃と想定した場合，提案手法では $n = 1$, $s = 15$, $h = 5$, $c = 20$ の時に本人拒否率と機械受入れ率を最小化できることを意味している．s の値を変化させた時の FRR と FAR の関係を図 6.5 に示す．

これより，この条件で提案手法による CAPTCHA を行った場合，FAR 及び EER はおよそ 2% となる事が予想される．同時に，実験 2 の応答時間の結果から，CAPTCHA に掛かる時間は平均で凡そ 308 秒となる．

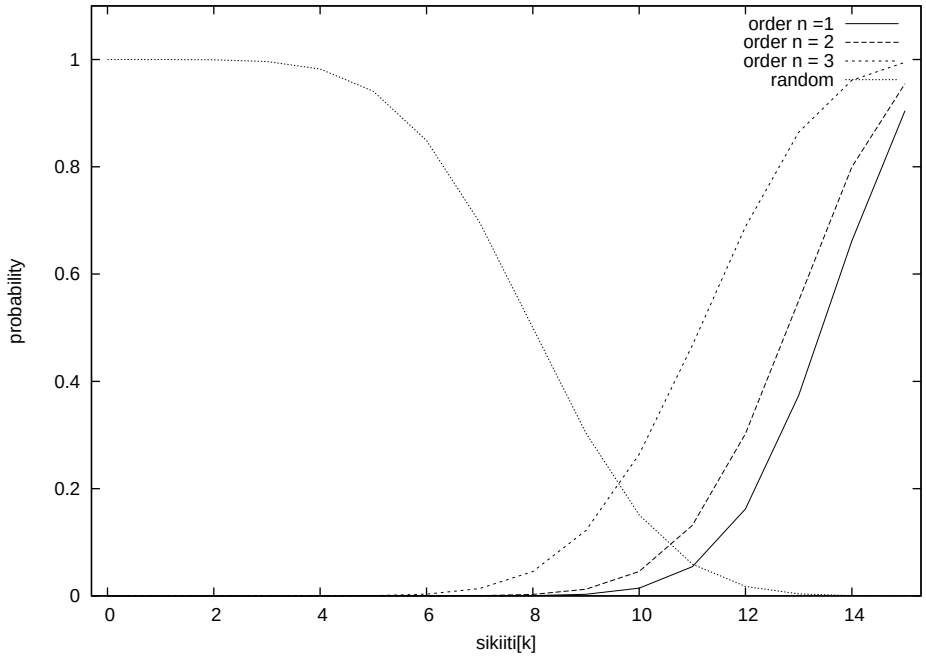

図 6.2: $c = 15$ の時の閾値 θ についての FRR と FAR

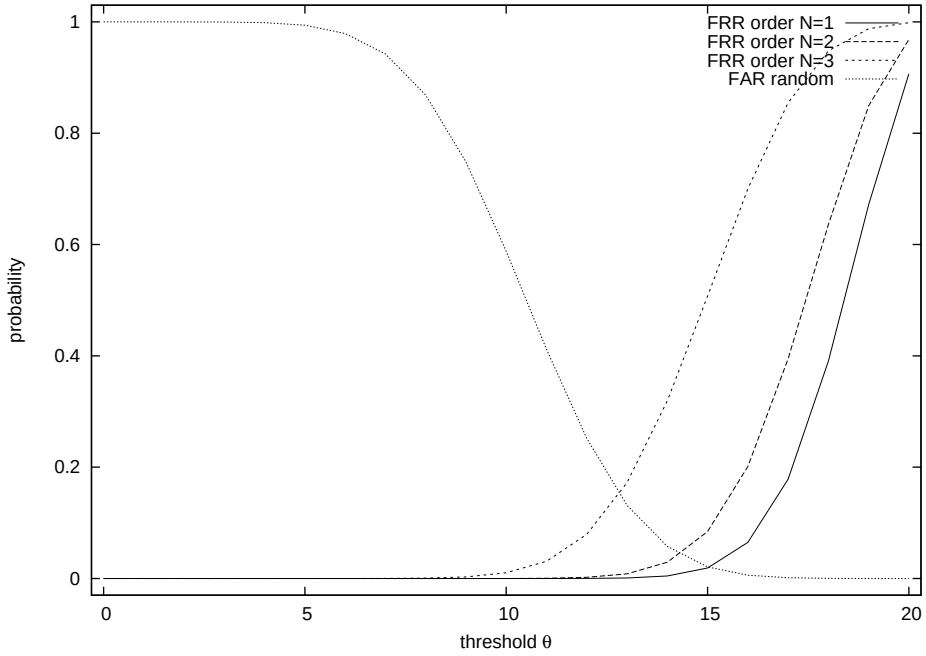

図 6.3: $c = 20$ の時の閾値 θ についての FRR と FAR

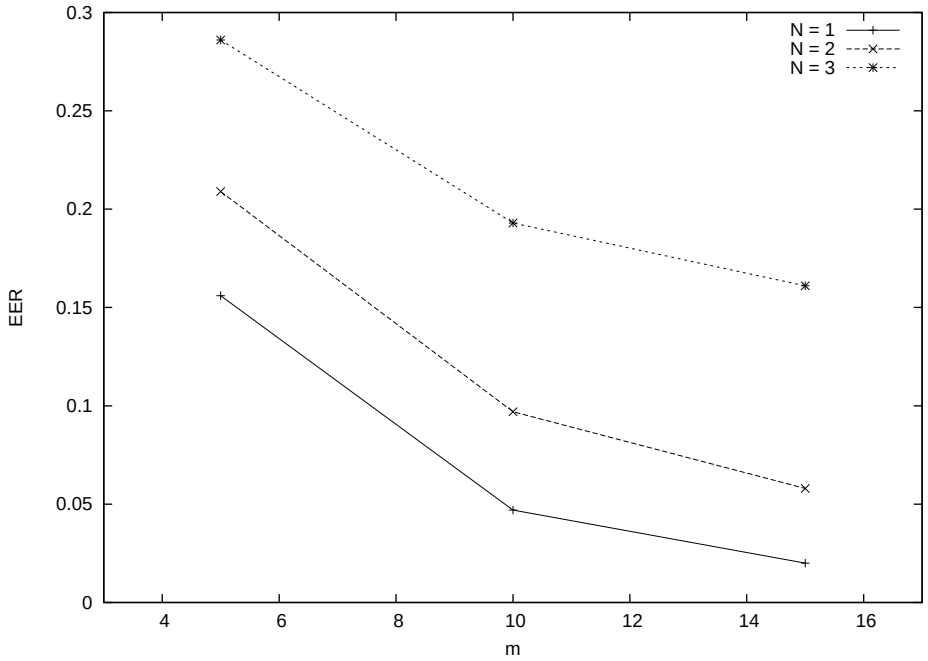

図 6.4: $n = 1, 2, 3$ の時の s についての EER

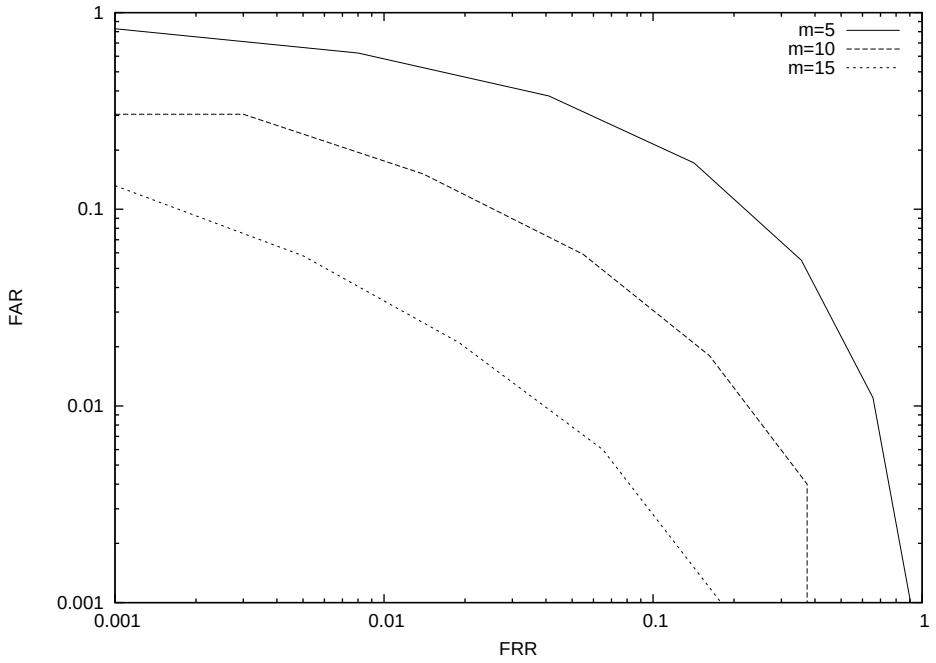

図 6.5: s の値を変化させた時の FRR についての FAR

6.2 解析 2

解析 2 では実験 3 の日本人留学生についての正答率を用いて，提案方式のリレーアタックへの耐性を評価する．提案方式に対するリレーアタックは，外国人低賃金労働者による攻撃であると想定する．

実験 3 から得られた $n = 1$ の時の解答の正答率は，条件付確率 $P(Y|X)$ として表 6.4 の様に与えられた．ここで基本解析と同様に，表 6.4 より，図 6.6 に示す外国人留学生の場合の FRR と FAR の関係を求める．

表 6.4: 留学生の $n = 1$, $s = h$ の時の条件付確率 $P = (Y|X)$

入力文書 \ 判別文書	$Y = H$	$Y = S$
$X = H$	0.50	0.50
$X = S$	0.20	0.80

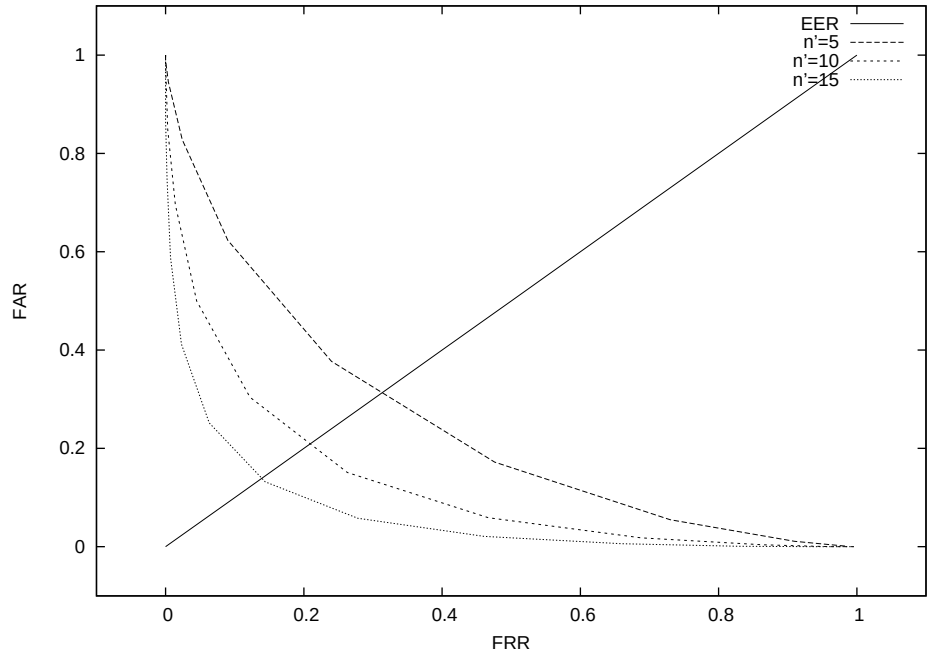

図 6.6: 留学生の場合の，s の値を変化させた時の FRR についての FAR

図**??**より，実験 3 の結果からも基本解析と同様に EER が求められる事が解る．基本解析で得られた EER が最小となる条件である，$n = 1$, $s = 15$, $h = 5$, $c = 20$ の時の FRR と FAR について，図 6.3 との比較を図 6.7 に，と図 6.5 との比較を図 6.8 にそれぞれ示す．
図 6.7，図 6.8 より，　提案手法に置いては日本人と外国人との間の精度に大きな差がある事が解る．

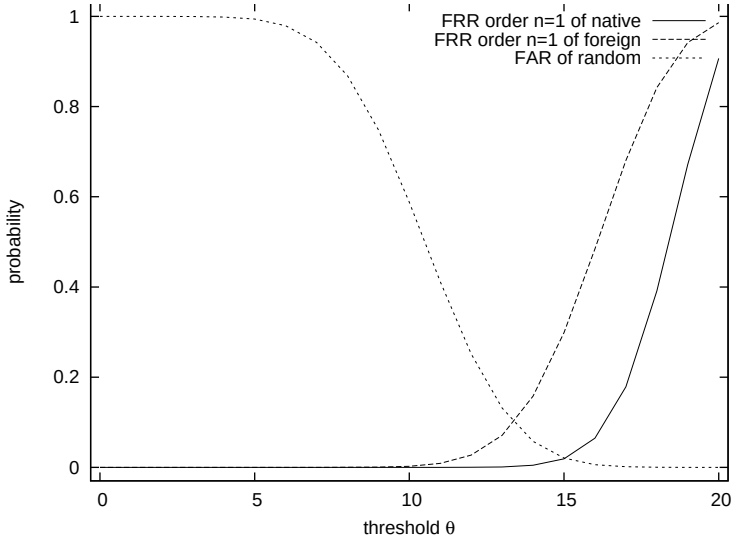

図 6.7: $c = 20$ の時の閾値 θ についての FRR と FAR の基本解析との比較

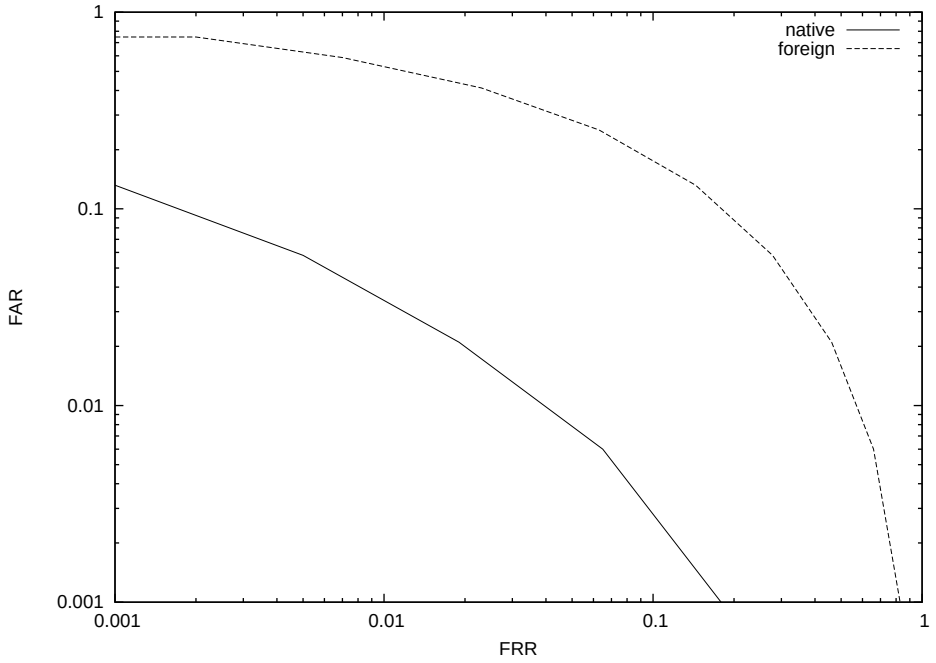

図 6.8: $n = 1$ の時の，FRR についての FAR の基本解析との比較

6.3 解析 3

解析 3 では，実験 4 で得られた正答率と応答時間を用いて，文章量を変化させた際の精度とパフォーマンスについて基本解析との比較を行う．実験 2 の結果と基本解析より，CAPTCHA として提示する文章量が 5 文の場合では CAPTHCA として提示する問題 1 題について，およそ 16 秒程の時間が掛かる事が解っている．その為，充分な精度を得る為には 308 秒の時間が掛かる事が予想されている．しかし提示する文章量を 1 文にして行った実験 4 の結果では，1 題について応答時間は平均して約 8 秒程になるという結果が得られている．従って，提示する文章量を 5 文から 1 文に減らす事で CAPTCHA に掛かる時間が半減出来ると考えられる．

実験 4 から得られた文章量が 1 文の時の $n = 1$ の時の解答の正答率は，条件付確率 $P(Y|X)$ として表 6.5 の様に与えられる．

表 6.5: 文章量が 1 文の時の $n = 1$ の時の条件付確率 $P = (Y|X)$

入力文書 \ 判別文書	$Y = H$	$Y = S$
$X = H$	0.91	0.09
$X = S$	0.27	0.73

表 6.5 を元に，提示する文章量が 1 文の場合について，基本解析と同様に EER が c の値が大きくなりすぎない範囲で最も低くなる条件を求めた．その結果，$n = 1$, $s = 5$, $h = 15$, $c = 20$, $\theta = 15$ の時に最も適切な値となる事が解った．提示する文章の量が 5 文の場合と 1 文の場合について，互いに最も精度が高くなる時の FRR についての FAR の比較を図 6.9 に示す．

以上の解析より，CAPTCHA として提示する文章量が 1 文の時，提案手法では $n = 1$, $s = 5$, $h = 15$, $c = 20$, $\theta = 15$ の時に最も精度が良くなり，$FAR = 4.3\%$ 及び $FRR = 2\%$, CAPTCHA に掛かる時間は実験 4 の応答時間の結果より平均 151.7 秒程度となる事が予想される．

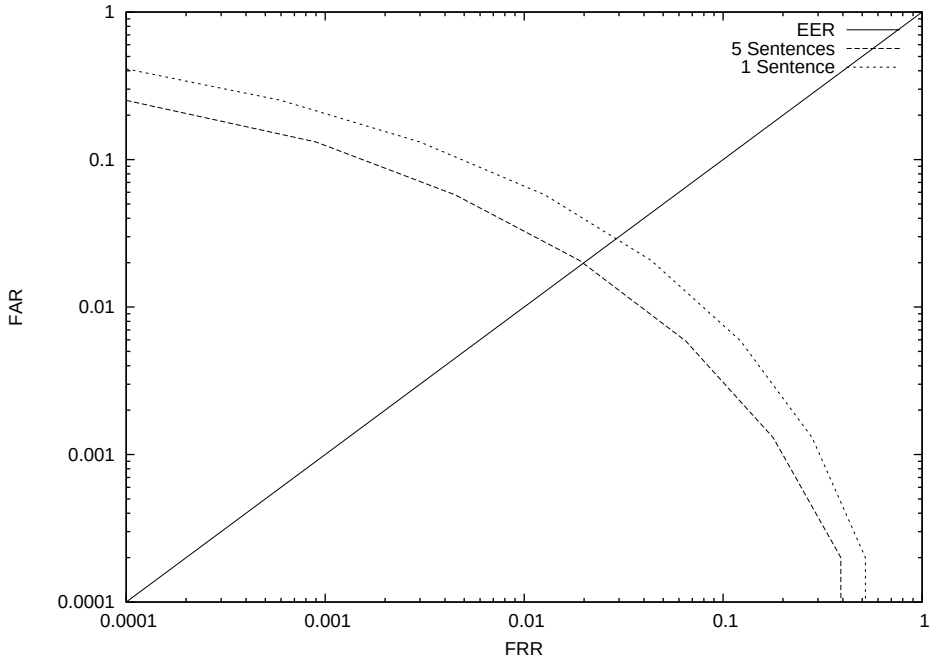

図 6.9: FRR についての FAR の提示する文章量による差の比較

6.4 解析 4

解析 4 では，実験 5 で行った様な文章校正による検出を用いた攻撃を行われた場合の提案手法の精度について検討する．提案方式について，一つの文について文章校正が行われる確率を $P(w)$ とする．実験 5 の結果より，MicroSoft Word による文章校正を 1 文ずつ行った際に文章構成が行われる確率は，条件付確率として以下の様に与えられた．

$$\begin{aligned} P(w|X=S) &= 0.24 \\ P(\overline{w}|X=H) &= 0 \end{aligned}$$

解析 3 より得られた最適な条件である $s=15, h=5$ で提案手法を行った場合，自然な文 H が出題される確率は $P(X=H)=0.75$，不自然な文 S が出題される確率は $P(X=S)=0.25$ となる．ここで，$P(w)$ は同時確率として以下の様に表せる．

$$P(w) = P(w, X=S) + P(w, X=H)$$

この同時確率は，同時確率の定義に従い，条件付き確率として以下の様に求められる．

$$P(w) = P(w|X=S)P(X=S) + P(w|X=H)P(X=H)$$

これを解くと，提案手法において 1 題あたり検出が行われる確率は $P(w)=0.06, P(\overline{w})=0.94$ と与えられた．この時，が行われた文の入力がスパムである確率は，条件付き確率で以下の様に求められる．

$$P(X=S|w) = \frac{P(w|X=S)P(X=S)}{P(w)}$$

同様に検出が行われなかった際の入力が自然な文である確率は以下の様になる．

$$P(X=H|\overline{w}) = \frac{P(\overline{w}|X=H)P(X=H)}{P(\overline{w})}$$

従って，機械は表 6.6 に示す文章校正を用いた検出によるスパム判定機を得る．この時，機械は提案手法に置いての s, h の割合を知らないと仮定し，機械は $P(X=S)=0.5$ 及び $P(X=H)=0.5$ として判定機を得るとする．

表 6.6: 文章校正を用いた検出によるスパム判定機

入力文書 \ 判定	$\overline{w}$	w
$X=H$	0.6	0
$X=S$	0.4	1.0

この判定機を用いた攻撃では，機械は文章校正による検出が行われれば必ず「不自然」であると解答し，そうで無い場合は 60% の確率で自然を，40% の確率で不自然を選択する．この「自然」「不自然」の選択は機械による出力がそれぞれ $Y=H$，$Y=S$ となる事を表す．

この判定機による判定が実際に成功する確率は，それぞれの同時確率で $P(Y = S, X = S)$ と $P(Y = H, X = H)$ の二種類であり，これらは，条件付き確率として以下の様に求められる．

$$\begin{aligned} P(Y = S|X = S) &= P(Y = H|w)P(w) + P(Y = H|\overline{w})P(\overline{w}) \\ P(Y = H|X = H) &= P(Y = H|w)P(w) + P(Y = H|\overline{w})P(\overline{w}) \end{aligned}$$

これらを基に，機械による攻撃の成功率は条件付確率 $P(Y|X)$ として表 6.7 の様に求めた．

表 6.7: 判定機を用いた機械による条件付確率 $P = (Y|X)$

入力文書 \ 判別文書	$Y = H$	$Y = S$
$X = H$	0.564	0.436
$X = S$	0.436	0.564

この機械による攻撃の成功率を $P_q w$ とし，この判定器を用いた機会受けれ率 FAR_w を以下の様に定義する．

$$FAR_w = \sum_{k=\theta}^{s} \binom{s}{k} P_q w^{s-k} (1 - P_q w)^k$$

以上の値を用いて，解析 3 との比較を行う．

解析 3 の結果より得られた最適な条件である，提案方式による CAPTCHA の問題として提示する文章量が 1 文の時，$n = 1, s = 5, h = 15, c = 20$ の場合に置いて，正解数 k の閾値 θ についての FAR_w，FAR と FRR を図 6.10 に，FRR についての FAR,FAR_w を図 6.11 にそれぞれ示す．

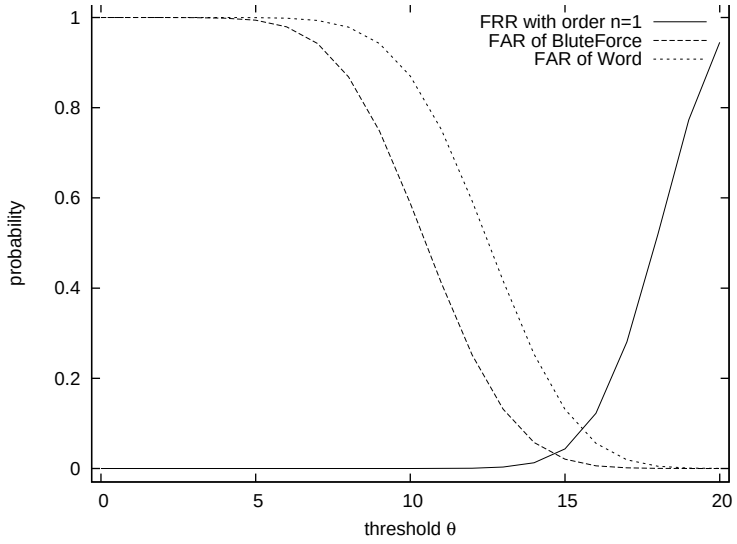

図 6.10: 正解数 k の閾値 θ についての FAR_w，FAR と FRR

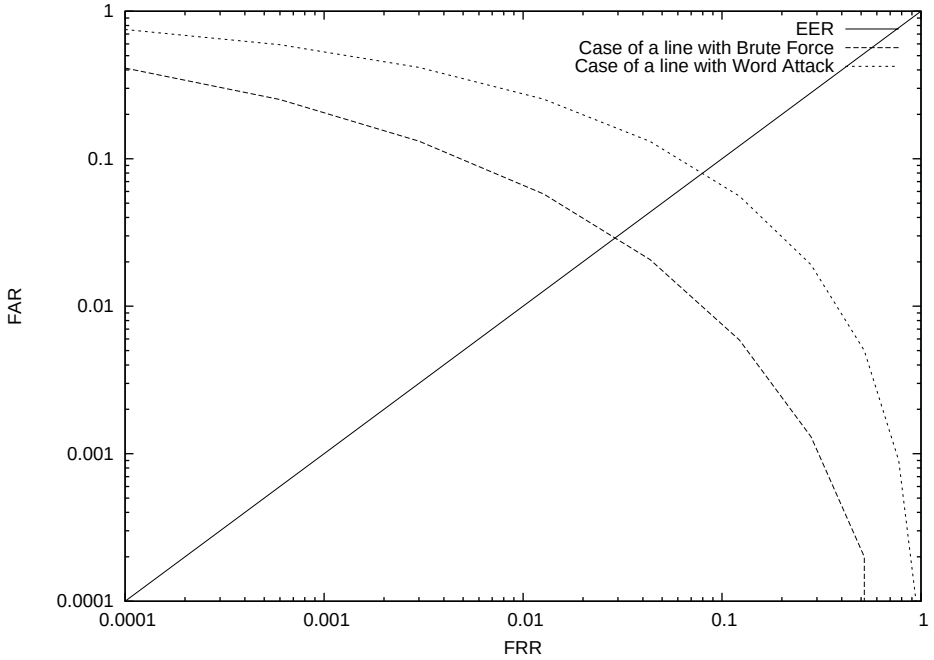

図 6.11: FRR についての FAR,FAR_w

以上の解析より，解析 3 と同様の $n = 1, s = 5, h = 15, c = 20$ の条件で提案方式を行い，機械により文章校正ツールを用いた検出が行われた時，EER は 8%程度となり，閾値 $\theta = 15$ の時 $FRR = 4.3\%, FAR_w = 13.1\%$，あるいは閾値 $\theta = 16$ の時 $FRR = 12.2\%$,$FARw = 5.6\%$ となる事が予想される．これらの関係を以下の表 6.8 に纏める．

表 6.8: 閾値 θ についての精度の比較

θ	FRR	FAR	FAR_w
15	0.434	0.020	0.131
16	0.122	0.006	0.057

6.5 解析による結論

精度の指標を機械受け入れ率及び人間拒否率値が同じとなる EER とする．基本解析では提示する文章量が 5 文の時，$n = 1, s = 15, h = 5, c = 20$ の時，EER は 2% と非常に良い精度で CAPTCHA が行える事が解った．しかし，従来手法の平均応答時間は 10 秒程度であるのに対し，実験 2 の結果より階数 $n = 1$ のワードサラダの平均応答時間は 13.3 秒，自然な文の平均応答時間は 21.07 秒であり，自然な文を 5 題，不自然な文を 15 題提示するこの条件では，CAPTCHA にかかる合計時間はおよそ 307.85 秒となると予想された．

解析 2 では閾値 θ を日本人の EER が最小となる $\theta = 15$ と設定した時，日本語を学んだ留学生であってもその FAR は 50%程度となる事が解った．リレーアタックで CAPTCHA を解く人間は，日本語を学んでいない事が予想されるため，提案手法は正規ユーザの母国以外の不正ユーザからのリレーアタックを相当の割合で防ぐ事が可能であると言える．

解析 3 では，問題として提示する文章の量を 5 文から 1 文に減らしても，最適な EER には 1%程度の違いしかない事を明らかにした．この精度は CAPTCHA として利用する分には充分であると考えられる．よって，パフォーマンスの向上も伺える事から，提案手法の CAPTCHA として提示する文章量は 1 文とする事が望ましい．機械による攻撃が総当り攻撃だと仮定した時，$n = 1, s = 5, h = 15, c = 20, \theta = 15$ の時に $EER = 3.7\%$ と最も精度が良くなり，$FAR = 4.3\%$，及び $FRR = 2\%$ の精度で CAPTCHA が行える事が期待できる．同時に，CAPTCHA に掛かる時間は実験 4 の応答時間の結果より平均 151.7 秒程度となる事が予想され，基本解析のおよそ 2 倍のパフォーマンスを得る事が出来る事が解った．

解析 4 では，機械が文章校正ツールによる攻撃を行うと仮定した場合について検討を行った．結果として，提案方式がこの手法の攻撃を受けた場合，EER は 8%となり，精度が半減してしまう事がわかった．

以上の解析より，機械による攻撃を総当たり攻撃と仮定した時，提案手法の CAPTCHA として適切なパラメータは，$s = 5, h = 15, c = 20, n = 1, \theta = 15$ である．この条件の時，

FRR 及び FAR はそれぞれ 4.3%, 2.0% となり，パフォーマンスは平均で凡そ 151.7 秒であると結論付けた．また，提案手法による CAPTCHA は，リレーアタックに対して耐性を持つ事を示した．しかし，文章の不自然さを機械が判定する事で機械受け入れ率を上げる事が出来る可能性があり，機械による攻撃については更なる対策が必要である事が伺える．

第7章
おわりに

7.1 結論

本論文では，マルコフ連鎖により合成された文章の不自然さを応用したCAPTCHAの提案を行い，ワードサラダの不自然さを評価する実験データからその性能を評価した．20題の文章中に階数$n = 1$のワードサラダ5題と自然な文15題という最も精度の良くなる条件下では，人間拒否率4.3%及び機械受入れ率2%の精度と，必要時間151.7秒のパフォーマンスで認証を行う事が可能である事を明らかにした．更に，提案方式がリレーアタックに対して耐性を持つ事と，日本語に限らず他言語にも適用可能である事を示した．

7.2 考察

実験で使った問題は自然な文章を合成しやすくなるようにコーパスの規模を5000文字から10000文字と非常に少ないものにしている．そのため自然なワードサラダも多く出力され，FARは高くなっていると思われる．また，提示する文章量1行にまで減らしたが，従来手法に比べユーザに掛かる負荷は依然としてとても高い．ワードサラダ合成に用いるコーパスをより大きくする事で，より不自然な文章が合成できることを期待できる．また，出現頻度の低い単語へのマルコフ連鎖の遷移確率を高くする事により，不自然な文章を合成できる確率が高まるものと思われる．しかしその場合，文として成立するかどうかは検討の余地がある．ワードサラダに充分な不自然さが与える事が出来れば，問題数を減らし，パフォーマンスを上げる事も可能である．

CAPTCHAとして適切な精度の値については，例えば総当り攻撃を想定した時のEERの値が10%程度となる閾値と問題数を設定したとする．その時，FARは10%であるので，機械は7回も攻撃を繰り返せば50%以上の確率でCAPTCHAを成功させてしまえる事になる．その為繰り返し攻撃による耐性を得る為にCAPTCHA試行回数に制限を設ける事は考えられる．試行回数制限を2回とすれば繰り返し攻撃によるCAPTCHA成功率は19%程度となるが，その場合FRRも同様に10%である為，同じ19%の確率で試行回数制限に掛かってしまう正規ユーザーも存在する事となる．そして，提案方式において$EER = 0.1$の条件で最もパフォーマンスが良い時，CAPTCHAに要する時間は105秒程である．正規ユーザー

が問題を読まずに適当に回答するのに必要な時間は，実験 2 の結果から，1 題あたり 0.2 秒程と得られている．つまり適当に回答を行った場合必要時間は 4 秒であり，それを 22 回も繰り返すと CAPTCHA 成功率は真面目に CAPTCHA を行った場合と同じ 90%となる．その際に必要となる処理時間は 88 秒と予想される為，こちらの方がパフォーマンスが良いと言う事になり兼ねない．こうなると人間の高度な認知処理を用いた CAPTCHA とは言い難くなってしまう．

また，解析 4 では機械が CAPTCHA の正解の割合を知らないと仮定したが，実際にはそれを知る事は容易に可能であると考えられる．その為，問題として提示する自然な文，不自然な文の割合は 1:1 である事が望ましいが，その時の提案手法の EER は 14%程度となる．無論，EER を下げる為に閾値や問題数の条件を変化させる事は可能だが，その場合パフォーマンスの低下は避けられない．これは，不自然な文を含む割合を変化させた時にも同様である．よって，提案手法の性能を上げる為には，より精度良く，短時間で人間にとって識別可能な問題を作成する事が必須であると考える．

7.3 今後の課題

より不自然さを保証した文章合成方法の検討と，提案方式の自動化を今後の課題とする．

参 考 文 献

[1] The Official CAPTCHA Site,
(http://www.captcha.net)

[2] J. Yan and A. S. E. Ahmad: Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms, 2007 Computer Security Applications Conference, pp. 279-291，2007.

[3] J. Elson，J. Douceur，J. Howell and J. Saul，Asirra: a CAPTCHA that exploit interest-aligned manual image categorization, 2007 ACM CSS, pp. 366-374, 2007.

[4] P. Golle: Machine Learning Attacks Against the ASIRRA CAPTCHA, 2008 ACM CSS, pp. 535-542 2008.

[5] 山本匠，J. D. Tygar, 西垣正勝：機械翻訳の違和感を用いた CAPTCHA の提案，情報処理学会研究報告, CSEC-46 No. 37，2009.

[6] 山本匠，J. D. Tygar, 西垣正勝：機械翻訳 CAPTCHA(その 2)，コンピュータセキュリティシンポジウム 2009 論文集，pp. 211-216 (2009.10)

[7] 鈴木 徳一郎，山本匠，西垣正勝：リレーアタックに耐性をもつ CAPTCHA の提案，情報処理学会研究報告, CSEC-48 No. 16, 2010.

[8] T. Larvergne, et al.,∶Detecting Fake Content with Relatine Entropy Scoring', CEVR, Vol.377, pp. 27-31, 2008.

[9] 森本，片瀬，山名：N-gram と離散型共起表現を用いたワードサラダ型スパム検出手法の提案, 情報処理学会研究報告, DBS-148, No.24, pp.1-8,2009.

[10] 鴨志田芳典，菊池浩明：マルコフチェインによるワードスパムの合成実験とその評価について，第 72 回情報処理学会全国大会，講演番号 2G-1，2010.

[11] MeCab，MeCab: Yet Another Part-of-Speech and Morphological Analyzer,
(http://mecab.sourceforge.net/)

謝辞

本研究を遂行するに辺り御指導，御鞭撻を賜りました，

東海大学情報通信学部通信ネットワーク工学科教授菊池浩明教授に最大の感謝を申し上げます．御迷惑ばかり掛けてしまい申し訳御座いません．

そして，度重なる御指導を頂いた情報理工学部情報科学科内田理准教授と中西祥八郎先生に多大なる感謝を申し上げます．不甲斐ない事ばかりで申し訳ありません．

また実験に御協力下さった方々と，三年間，共に学生生活を楽しんだ研究室の仲間達に感謝を申し上げます．

最後に，東海大学は菊池研究室に来る事が出来た天命に感謝の意を表すると共に，これを謝辞とさせて頂きます．