機械翻訳 CAPTCHA（その 2）

山本 匠[1,3]　J. D. Tygar[2]　西垣 正勝[1,4]
[1] 静岡大学創造科学技術大学院
[2] Computer Science Division, University of California, Berkeley
[3] 日本学術振興会特別研究員（DC）　[4] 科学技術振興機構，CREST

あらまし　近年，既存の CAPTCHA における脆弱性が多くの研究者によって指摘されており，人間の「より高度な知識処理」に基づいた新たな CAPTCHA の導入が強く望まれる．そこで著者らは既に，「より高度な認知処理」の 1 つとして「違和感の判別」をチューリングテストに用いることで，人間には容易で機械には困難な CAPTCHA を提案している．さらに，その実現の第一歩として，機械翻訳により生成された文章が有する違和感に注目した SS-CAPTCHA を実装し評価を行った．しかし，既存研究では SS-CAPTCHA における攻撃耐性についての検討が不十分であったため，提案方式の有効性が不明確であった．そこで，本稿では，SS-CAPTCHA に対するいくつかの攻撃を考え，それらへの耐性および対策について調査・検討を行う．

CAPTCHA using Strangeness in Machine Translation (part2)

Takumi Yamamoto[1,3]　J. D. Tygar[2]　and　Masakatsu Nishigaki[1,4]
[1] Graduate School of Science and Technology, Shizuoka University
[2] Computer Science Division, University of California, Berkeley
[3] Research Fellow of the Japan Society for the Promotion of Science (DC)
[4] Japan Science Technology and Agency, CREST

Abstract　Recently as many researchers have reported, the conventional CAPTCHA could be defeated by recent malwares since the ability of PCs get closer to that of human. Therefore CAPTCHA should be based on an even more advanced human cognitive processing ability. To realize a new CAPTCHA, we have proposed to use a human ability to recognize “strangeness”. As an example, we focused on strangeness in machine translated sentences, and studied a SS-CAPTCHA which detects malwares by checking if a user can distinguish natural sentences created by human from machine translated sentences. However, previous study has not mentioned possible attacks against SS-CAPTCHA and the effectiveness of SS-CAPTCHA have yet to be evaluated. Thus in this paper, we describe several possible attacks against SS-CAPTCHA and discuss their countermeasures through basic experiment.

1. はじめに

無料 Web メールやブログなどのインターネットにおける Web サービス提供サイトに対し，機械（マルウェア等の自動プログラム）を使って，大量にアカウントを不正取得する，多数のブログサイトにスパム記事を不正投稿する，大量に不正なサービス利用要求を行うなどのいわゆる DoS（Denial of Service，サービス不能）攻撃が定常的に頻発している．この問題に対してチューリングテスト（マルウェア等の悪意のある自動プログラムと正規のユーザ（人間）とを識別するためテスト）の有用性が高まっており，現在，CMU の研究者によって開発された CAPTCHA [1]と呼ばれる方式が広く利用されている．

CAPTCHA の基本形態は，歪曲やノイズが付加された文字列画像を Web ページに提示し，閲覧者がその文字を判読できるか否かを試すものである．この方式の CAPTCHA の例を図 1 に示す．しかし，近年，既存の CAPTCHA における脆弱性が多くの研究者によって指摘されている[2]．例えば，文字列の判読能力を試す CAPTCHA においては，すでに高機能な OCR（自動文字読取）機能を備えるマルウェアが出回るようになっている[3].

図 1　Google で使用されている CAPTCHA

文字列に加える変形やノイズを大きくすることによってマルウェアを排除する確率を向上させることはできるが，そのような文字は人間にとっても難読度が高まるため，人間の正答率まで低下させてしまう．この問題に対し，人間の「より高度な認知処理」を利用して CAPTCHA を強化する方法が検討されてきた．その代表的なものとして Asirra [4]がある．Asirra では，複数の動物の絵を表示し，その中から特定の動物の絵を選ばせる．例えば「猫を選べ」という質問に対し，猫の絵を正しくすべて選択することが

できれば人間であるとして判定する．「絵の意味を理解する」ことは人間の高度な認知メカニズムの一つであり，現在のレベルのマルウェアによる解答は非常に難しいと考えられていた．しかし，最近になって，Asirra を破る自動プログラムに関する研究報告がなされ，研究者の間に衝撃が走った[5]．

マルウェアの能力の向上は留まるところを知らない．マルウェアがいかに高度になろうともマルウェアによる解答が依然として困難な，人間の「より高度な認知処理」に基づいた新たな CAPTCHA の導入が強く望まれる．

そこで著者らは既に「より高度な認知処理」の１つとして「違和感の判別」をチューリングテストに用いることで，人間には容易で機械には困難な CAPTCHA を提案している[6]．さらに，その実現の第一歩として，機械翻訳により生成された文章が有する違和感に注目した SS-CAPTCHA (CAPTCHA using Strangeness in Sentences)を実装し評価を行っている．しかし，既存研究では SS-CAPTCHA における攻撃耐性についての検討が不十分であったため，提案方式の有効性が不明確であった．そこで，本稿では，SS-CAPTCHA に対するいくつかの攻撃を考え，それらへの耐性および対策について調査・検討を行う．

2. SS-CAPTCHA

本章では，SS-CAPTCHA のコンセプトと CAPTCHA に用いる文章の収集方法について簡潔に述べる．

2.1 コンセプト

SS-CAPTCHA とは，Web ページ訪問者に，人間が作成した自然な文章（以降，NS (Natural Sentence) または自然文章と呼ぶ）と機械翻訳により生成された文章（以降，GS (Garbage Sentence) または機械翻訳文章と呼ぶ）とを複数提示し，訪問者に NS を選択してもらうことで，訪問者が人間か機械であるかを判定する CAPTCHA である．

機械翻訳技術は目覚ましい進歩を遂げてきた．しかし，他言語の文章を機械翻訳にかけた日本語は，日本人にとっては依然として違和感を覚えるものがあり，自然な文章を自動的に作り出すことは非常に難しい技術の内の一つであると言える．これは，機械にとって自然言語を完全に解釈することが非常に困難であるという証拠に他ならない☆．

すなわち，現在の技術レベルを仮定した場合，機械が GS と NS との些細な違いを見つけることは不可能に近いといえる．一方，人間であれば，通常，違和感のある不自然な母国語文章を簡単に見つけることができる．図 2 に SS-CAPTCHA の概観を示す．

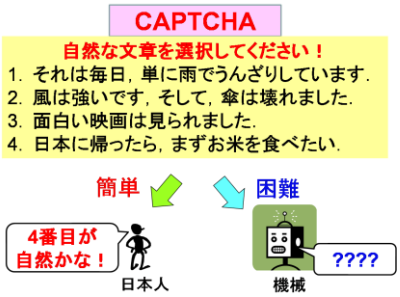

図２　SS-CAPTCHA の概観

2.2 文章の収集

● 機械翻訳文章（GS）の収集

GS は機械翻訳プログラムを用いて，母国語以外の任意の自然な言語を母国語に翻訳することによって生成される．母国語を別の言語に翻訳した上で元の言語に再翻訳したり（例，日本語→英語→日本語），複数の言語間で機械翻訳を幾重にも繰り返して（例，英語→イタリア語→英語→フランス語→日本語）も良い．同じ言語に対する異なる機械翻訳プログラムを組み合わせて使用しても構わない．

● 自然文章（NS）の収集

Web 上には人間が作成した文章が無数に存在するが，これらを SS-CAPTCHA の NS として利用することは難しい．なぜなら，そのような文章は機械も Web 検索エンジンを使って NS であるか否かを効率よく判定することが可能だからである．

そのため，SS-CAPTCHA で用いる NS は Web 上には存在しない文章が望まれる．そのような文章は，日々新しく作られる新聞，雑誌，書籍等から抽出することが可能であると考えられる．また，日々新しい内容が報道されるニュース番組内でアナウンサーが話す自然で正しい文章を，音声認識技術を用いることで，自動的に収集することも可能になると考えられる．

● 文章の評価

抽出した NS の全てが人間にとって違和感のない文章で，機械翻訳により生成された GS の全てが人間にとって違和感のある文章であるとは限らない．ときに，違和感のある NS が得られたり，違和感のない GS が得られたりする場合もあるだろう．このような文章をそのまま NS と GS として SS-CAPTCHA に導入することは，人間を惑わす大きな要因となる．

そこで，SS-CAPTCHA では，得られたばかりの NS や GS をすぐに CAPTCHA チャレンジとしての文章としては用いず，しばらくの間，「試験段階の文章」として扱う．試験段階の NS および GS は，その時点ですでに CAPTCHA チャレンジ用文章として確定している複数の

☆ もし，機械翻訳文章に対して，人間が覚えるような違和感を機械も認識することができたなら，現在の機械翻訳プログラムは自分の翻訳結果をセルフチェックすることによってより適正な文章を出力することができるようになっていて不思議ではない．

NS と GS とともに CAPTCHA 画面に提示される．

表示された文章のうち，チャレンジ用文章として確定している NS および GS に対する訪問者の解答から，訪問者が人間であるか機械であるかを切り分けることができる．訪問者が人間であると判断された場合には，表示された文章のうち，試験段階の NS および GS に対する訪問者の解答を検査する．訪問者が違和感がないと解答した試験段階の文章は，NS である可能性が高いと判断され，そうでない文章は GS である可能性が高いと判断される．このようにして，確実に自然な文章の集合と確実に違和感のある文章の集合を構成していく．提案方式の詳細については既存研究[6]を参照されたい．

3. SS-CAPTCHA への攻撃とその対策

本章では既存研究[6]では議論されなかった，SS-CAPTCHA に対して効果的に実行可能な攻撃方法について紹介し，その対策について議論する．

3.1 翻訳の収束性に基づく攻撃

3.1.1 翻訳の収束性

ある文章 S を別の言語に機械翻訳した上で，元の言語に再翻訳する変換を関数 F として表現する．ある文章 S_1 に関数 F を適応して得られた機械翻訳文章 $S_2 = F(S_1)$に対し，再び関数 F を適応して得られた文章を $S_3 = F(S_2)$と表現する．この作業を繰り返していくと，元の文章 S_1 に依存する一定の繰返し回数 C (以降　収束回数と呼ぶ) で S_C と S_{C+1} が等しくなる（$S_C = S_{C+1}$）場合がある．さらに収束回数 C には偏りがあり，S_1 が NS のときは C=2 となるケースが多いであろう．そこで，収束回数の傾向について調査したところ，図 3 のような収束回数の分布が見られた．

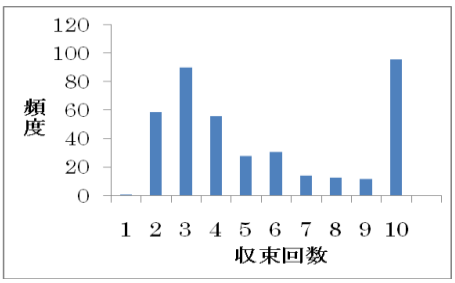

図 3　収束回数の傾向

図 3 は，あらかじめ用意した NS に関数 F を適用し，文章ごと収束回数 C を求め，その傾向をヒストグラムに表したものである．本調査で用いた NS は，新聞[7]と書籍[8]から適当な長さ(11～58 字)の文章をそれぞれ 200 個抽出して得られたものである．また今回は，翻訳前および翻訳後の言語をそれぞれ日本語と英語に絞り，エキサイト翻訳[9]を用いて調査した．なお，収束回数が増加するにつれ分析に非常に時間を要するようになるため，本調査では，収束回数の上限を 10 に固定し，10 回変換した時点で収束しない文章の収束回数は 10 とし，その時点で変換を停止した．

図 3 より，収束回数 C=10 のときを除けば，NS から 2~4 回の変換（関数 F は，日本語文章を英語に翻訳した上で，日本語に再翻訳する変換を 1 回と数えることに注意）で，変換結果が収束することが多いことがわかる．また，SS-CAPTCHA では自然文章 NS の機械翻訳によって機械翻訳文章 GS を生成する (GS=F(NS)) ことを考慮すれば，GS（$=F(S_1)$）を起点とした変換の収束回数は，NS（$=S_1$）を起点とした収束回数 C から 1 を引いたものとなることがわかる．よって，GS の変換については，1~3 回の変換で結果が収束する傾向にあるといえる．

この特性は，CAPTCHA 画面に表示される複数の文章（NS と GS）に対し，それぞれの文章を何度も機械翻訳にかけ，変換の収束回数 C が 1 の文章，または，相対的に小さい文章は GS の可能性が高いという推測方法を攻撃者に与えてしまう．以降，本攻撃のことを収束解析攻撃とよび，次節ではその対策を議論する．

3.1.2 収束解析攻撃への対策

● NS と GS の収束回数の調整（分布調整法）

収束解析攻撃は，CAPTCHA 画面に表示される NS と GS の収束回数（または収束回数の分布）が異なっているために起こりうる．すなわち，画面に表示されている NS と GS の収束回数（または収束回数の分布）が同程度であれば，収束解析攻撃による推測は困難になると考えられる．

そこで対策として，画面に表示される複数の GS の収束回数の平均値(m)および標準偏差(σ) が，画面に表示される複数の NS のそれらと同程度になるように，NS と GS を選択する．以降，　本対策手法のことを分布調整法と呼ぶ．

● 正確な収束結果の取得の困難化（カタカナ法）

収束解析攻撃は，CAPTCHA 画面に表示される文章の収束回数を，機械翻訳ツールを用いて機械的に計測可能なために起こりうるとも考えられる．つまり，機械的な収束回数の計測が困難であれば，収束解析攻撃の脅威を低下させることができると考えられる．

そこで対策として，画面に表示する文章を全てカタカナに変換して表示する．機械にとっては，カタカナのみから構成される文章（以降　カタカナ文章と呼ぶ）を，漢字と平仮名が混ざった正しい文章に変換することは難しい．すなわち，カタカナ文章の字句解析・意味解析は機械にとって困難な作業の一つである．一般的な日英機械翻訳プログラムでは，カタカナ表記の語句についても日本語辞書に登録されてはいる[10]が，カタカナ文章の解析自体が難しいため，カタカナ文章の入力に対しては，全文が単純にローマ字表記に変換された文章（以降，ローマ字文章と呼ぶ）

が出力されるという仕様になっていることが多い．同様の理由で，日英機械翻訳プログラムにローマ字文章の日本語を入力した場合は，入力されたローマ字文章が無変換でそのまま出力される傾向にある．また，ローマ字文章を英日機械翻訳プログラムに入力して再翻訳した場合も，入力されたローマ字文章が無変換でそのまま出力されるという仕様になっていることが一般的である．

以上より，カタカナ文章を日英機械翻訳した上で英日機械翻訳により再翻訳した場合は，カタカナ文章が単純にローマ字文章に変換された文章が得られる．そして，そのローマ字文章をもう一度，日英機械翻訳した上で英日機械翻訳した場合には，そのローマ字文章がそのまま無変換で出力される．そのため CAPTCHA の文章をカタカナにすることで，文章が NS であるか GS であるかに関係なく，ほとんどすべての文章の収束回数を 2 にすることができ，収束解析攻撃の脅威が低下すると考えられる．

ただし，本対策はカタカナという日本語特有の表記方法を用いているため，他の言語に適応することはできない．またカタカナだけの文章は人間にとっても単語の意味の推定が困難になり，CAPTCHA レスポンスに対するユーザの負担が大きくなる可能性がある．以降，本手法のことをカタカナ法と呼ぶ．

3.1.3 対策導入時の利便性に関する実験調査

3.1.2 節で議論した 2 種類の対策の導入が，機械ではなく人間の負荷増大につながらないかを確認する．本実験で用いた実験システムは，既存研究[6]の 4.2 節で使われた実験システムを元にしている．既存研究における実験システムとの差異は，本稿で提案した対策（分布調整法，カタカナ法）が導入されている他に，実験で利用する文章として 3.1.2 節で用意した文章（新聞や書籍から収集した文章）を用いている点である（既存研究では，被験者によって作られた文章が用いられていた）．

● 実験方法

以下に示す各 200 個の文章からなる 4 種類の自然文章の集合 NS_i (i=1~4)と， NS_1，NS_2 からそれぞれ生成された機械翻訳文章の集合 GS_1，GS_2，および，GS_1 と GS_2 をそれぞれカタカナ表記とした文章の集合 GS_3，GS_4 を用いて実験を行う．NS_i と GS_i(i=1,2)が分布調整法，NS_i と GS_i(i=3,4)がカタカナ法の有効性の確認に用いられる．

NS_1：新聞[7]から抽出した自然文章の集合

NS_2：書籍[8]から抽出した自然文章の集合

NS_3：NS_1 をカタカナに変換した自然文章の集合

NS_4：NS_2 をカタカナに変換した自然文章の集合

また，本実験では，既存研究で提案されている SS-CAPTCHA のモデル[6]が適切に機能することを想定し，不適切な GS☆をあらかじめ除外している．その結果，GS_1(GS_3)と GS_2(GS_4)中の文章の数はそれぞれ 155 個と 139 個になった．

以下に本実験の手順を示す．実験は，4 種類(i=1~4)のそれぞれの文章集合ごとに行われる．

1) システムは NS_i (i=1~4)の中からランダムに 5 個の自然文章 NS を選択する． そして 5 個の NS の収束回数の平均値(M_{NS})と標準偏差(σ_{NS})を求める．
2) システムは GS_i (i=1~4)の中からランダムに 10 個の機械翻訳文章 GS を選択する．このとき，i=1 または i=2 のときのみ，選択された 10 個の GS の収束回数の平均(M_{GS})と標準偏差(σ_{GS})がそれぞれ M_{NS} と σ_{NS} と同程度($|M_{NS} - M_{GS}| \leqq \theta_M$ かつ $|\sigma_{NS} - \sigma_{GS}| \leqq \theta_\sigma$)になるように選択する．なお， 1)で既に選択されている NS から生成された GS は選択しない．本実験では，$\theta_M = 0.5$， $\theta_\sigma = 0.5$ とした．
3) システムは選択した計 15 個の文章をランダムな順番で画面に並べる．
4) 被験者は与えられた 15 個の文章の中から違和感の無い文章を 5 個選択する．
5) 15 個の文章の中か 5 個の NS を正しく選択できたら，認証成功となる．

被験者には手順 1)~5)を 4 種類(i=1~4)の文章集合につき計 5 回ずつ実行してもらう．すなわち被験者は，15 個の文章の中から違和感の無い文章を 5 個選択するという作業を計 20 回実行する．1 回の選択作業の度に文章集合(i=1~4)がランダムに選択される．本実験の被験者は本学情報学部学生 6 名である．

● 実験結果

実験結果を表 1 に示す．表中，「解答時間」は 15 個の文章の中から 5 個の文章を選択し終えるまでに要した時間の平均である．「成功率」は CAPTCHA の解答に成功した割合を閾値 θ ごと示したものである．ここで，「閾値 θ」とは，画面に表示されている 5 個の NS のうち，選択失敗を許容する数（閾値）を意味する．

実験結果より，全体的に CAPTCHA の解答にかなり長い時間を要していることが見て取れる．既存研究[6]における解答時間 60~70 秒程度であることから，本実験で導入し

☆ 以下の(a), (b)が SS-CAPTCHA には不適切な GS である．(a) 翻訳後の日本語文章中に翻訳前の日本語には含まれていない記号やアルファベットが含まれている文章．(b) 違和感のない GS．(a) は翻訳プログラムで用いられる辞書に存在しない単語（または表現）が，翻訳前の文章に含まれていることが原因で起こる．(a)に該当する文章をそのまま利用することは機械に，当該文章が GS であることを教えるヒントになってしまう．(b)は人間に適切に違和感を与えることができない文章であり，人間を惑わす要因になりうる．

た対策がユーザの負荷を増加させている可能性がある．特にカタカナ文章(i=3,4)を使った CAPTCHA の解答時間は非常に長く，また，成功率も低い．被験者からの聞き取り調査では，カタカナ法は，「文字を 1 つずつ読んで，それぞれがどのような単語を構成しているのかを常に考えながら文章を読み解いていく必要があり，非常に大変であった」という意見が得られている．一方で，分布調整法は「出版物の文章（特に新聞）には難しい表現や単語が多くあり，目を通すのに苦労した」いった意見が多かった．これは分布調整法における負荷はその対策によって引き起こされたものではなく，今回用いた文章が原因であったと考えることができる．そのため今後は，ユーザの年齢，趣味嗜好に応じて利用する文章を選ぶなどの工夫も必要あろう．

表 1　実験結果

		文章の種類			
		NS_1	NS_2	NS_3	NS_4
解答時間[秒]		106.1	87.5	284.6	209.8
成功率[%]	$\theta=0$	63.3	56.7	33.3	60.0
	$\theta=1$	100.0	96.7	83.3	90.0
	$\theta=2$	100.0	100.0	96.7	100.0
	$\theta=3$	100.0	100.0	100.0	100.0

また，被験者からは「認証画面中の文章の数が多い」という意見も多く得られている．そこで，比較のために，NS または GS をランダムに 1 つ画面に表示し，表示された文章が NS なのか GS なのかをユーザに解答してもらうという非常にシンプルな 2 択方式☆を用いた実験を行い，解答時間と成功率を計測した．ただし，NS は NS_1 と NS_2 からそれぞれランダムに 5 個ずつ計 10 個の文章を，GS は GS_1 と GS_2 からそれぞれランダムに 5 個ずつ計 10 個の文章を選択し，被験者には 2 択の選択を計 20 回繰り返してもらった．その際，分布調整法を用い それぞれ 10 個の文章の NS と GS の収束回数の平均と分布が同程度になるよう調整した．実験結果を表 2 に示す．

表 2　2 択方式おける調査

	文章の種類	
	NS_1	NS_2
解答時間[秒]	7.4	6.7
成功率[%]	90.0	100.0

新聞(NS_1)に比べ書籍(NS_2)を用いた方式の結果が優れていることが見て取れる．また，表 1 の閾値 $\theta=0$ のときの結果と照らし合わせるために，2 択方式を 11 回または 12 回連続して繰り返した際☆の解答時間を試算☆☆☆すると，74~89 秒程度になることがわかり，シンプルな 2 択方式を複数回連続で繰り返した方が，人間の負荷が少ないとも考えることもできる．ただし，本調査は表 1 の実験が終了した後に，同じ文章集合を用いて同被験者に対して実施したものであり，学習効果や順序効果を加味していないため，信頼性に欠ける結果ではある．しかし，被験者からの聞き取り調査では「表 1 の実験で用いた方式よりも 2 択方式を十数回繰り返した方が簡単そう」といった意見が多く聞かれたことから，今後は文章の提示方法についても考慮しながら，分布調整法およびカタカナ法の負荷低減の工夫を検討していく予定である．

3.2 Web 検索結果の頻度解析による攻撃

3.2.1　NS と GS との検索結果の違い

Web 上には，人間が作った自然な文章（NS）が多数存在する．一方，機械翻訳により得られた不自然な文章（GS）が Web 上に掲載されることは特殊であると考えてよいだろう．すなわち，Web 検索エンジンによる文章の検索結果を比較することで，認証画面に表示されている文章のうちどの文章が NS なのか GS なのかを切り分けることが可能かもしれない．以降，本攻撃のことを検索攻撃と呼ぶ．

既存研究では，SS-CAPTCHA で用いる NS はあらかじめ，検索エンジンの全文一致検索で検索ができなかったもの（検索ヒット数が 0 件のもの）のみを利用することにしている[6]．しかし，たとえ，全文一致検索により当該文章が検索できなくとも，文章をフレーズごと切り分け，検索結果を比較していけば，NS の検索結果と GS の検索結果の間に差異が出てくるかもしれない．例えば，以下の 2 組の NS と GS のペアを見ていただきたい．

NS_a：私は今日中に提出しなければならない論文の執筆に追われています．

GS_a：私は，今日の終わりまでに提出されるべきである論文を書くことによって，追いかけられます．

NS_b：私はエアコンの風があまり好きではないので，窓を大きく開けて自然の風が通るようにしています．

GS_b：私がエアコンの風があまり好きでないので，窓を大いに開けて，自然の風に通らされます．

4 つの文章とも，全文を完全一致検索（Google 検索における引用符「“」「”」を用いた検索）によって検索した場

☆ 認証画面に NS または GS が 1 文表示され，ユーザはその文章に対して「違和感を覚えるか」，または「違和感を覚えないか」を回答する．

☆☆ 閾値 $\theta=0$ のとき，15 個の文章の中から 5 個の文章を選択する際の組み合わせ総数は ${}_{15}C_5=3003$ となる．一方，2 択方式を 11 回または 12 回連続して繰り返した際の組み合わせ総数はそれぞれ $2^{11}=2048$，$2^{12}=4096$ である．

☆☆☆ 連続して繰り返される 2 択それぞれの事象は，互いにほぼ独立であり，「2 択方式を n 回連続して繰返した際の解答時間」は「表 2 の解答時間」×「2 択の繰返し回数(n)」で求められる．

合，検索ヒット数が 0 件となる文章である．しかし，文章の中から特徴的なフレーズを抽出した場合，そのフレーズの完全一致検索の結果が NS と GS とで異なることが往々にして起こりえる．例えば，GS_aおよびGS_bの中の直観的に違和感を覚えるフレーズとして「提出されるべきである論文」，「窓を大いに開けて」に着目し，NS_aとNS_bの中から当該フレーズに対応するフレーズ（「提出しなければならない論文」，「窓を大きく開けて」）を抽出した場合，各フレーズを完全一致検索によって検索してみると，NS から抽出したフレーズは検索できるが，GS から抽出したフレーズは検索できないことが容易に確かめられる．このように，違和感のあるフレーズのペアを抽出し，完全一致検索の結果を比較することで，NS と GS を区別することができる可能性がある．

3.2.2 検索攻撃の可能性について

前節で述べたように，GS の中から違和感のあるフレーズを抽出し，それに対応するフレーズを NS からも抽出し，2 つのフレーズ（GS と NS のフレーズのペア）の完全一致検索結果を比較することで，検索攻撃が成功する可能性があると考えられる．

しかし，SS-CAPTCHA[6]では，ある GS とその元となる NS を同時に認証画面に表示することはしないという方法を採っている．よって，一つの認証画面に表示される文章の中には GS と NS のフレーズのペアが存在せず，検索結果を比較するということ自体が不可能である．

表 3　フレーズの検索結果の例 (2009 年 9 月 確認)

フレーズ	検索結果	フレーズの切り出し元
今日中に提出しなければ	15,300	NS_a
あまり好きでないので	99,100	GS_b
窓を大きく開けて	27,500	NS_b
論文を書くことによって	22,000	GS_a
自然の風が通るように	234	NS_b
今日の終わりまでに	73,000	GS_a

また，一般的に人間の直観的（感覚的）な処理を機械的に実装することは困難であり，違和感を覚えるフレーズを機械に抽出させることは容易なことではない．表 3 に NS と GS から適当に切り出したフレーズの完全一致検索結果（検索ヒット数）の例を示したが，フレーズを検索語とした場合の検索結果は，フレーズ（単語や表現）によって大きく異なる．よって，フレーズの検索結果から当該フレーズの違和感を推定するようなことも難しいであろう．文章中から任意のフレーズを総当り的に抽出して検索攻撃を行う場合には，一回あたりの検索攻撃にて検索エンジンを多数回利用する必要があるため，効率的な攻撃であるとはいいにくい．

以上を踏まえると，SS-CAPTCHA への検索攻撃の実行は現実的ではないと考えられる．

4. まとめと今後の課題

本稿では，機械翻訳により生成された文章が有する違和感に注目した SS-CAPTCHA に対する攻撃方法として，収束解析攻撃と検索攻撃について考察した．検索攻撃は現実的ではないという考察結果に至ったが，収束解析攻撃に対しては，本稿で提案した対策は人間の認証負荷を高める傾向にあり，さらなる改良が必要であることがわかった．今後，SS-CAPTCHA における文章の提示方法や，文章の抽出方法等を工夫することで，認証成功率の向上とユーザの認証負荷の低減を検討していきたい．

謝辞　富士通研究所 鳥居悟様，横浜国立大学 大石和臣様，KDDI 研究所　竹森敬祐様，情報通信研究機構 中里純二様に本方式への攻撃に関しての助言を頂いた．ここに深く謝意を表する．また，本研究は科研費（No.20-6290）の研究助成を受けている．

参考文献

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

[2] PWNtcha-Captcha Decoder, http://caca.zoy.org/wiki/PWNtcha

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

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

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

[6] 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

[7] 静岡新聞，2009 年 8 月

[8] 渋谷昌三：「しぐさ」を見れば心の 9 割がわかる！，王様文庫，三笠書房

[9] エキサイト翻訳, http://www.excite.co.jp/world/

[10] 松尾義博，畑山満美子，池原悟：英語辞書と英文法を用いたカタカナ標記語の翻訳，情報処理学会，第 53 回全国大会，Vol.2, pp.65-66, (1996 秋)