ipsjstyle-ms2009.dot (V2.0)

機械翻訳の違和感を用いた CAPTCHA の提案

山本匠[†,‡]　J. D. Tygar[††]　西垣　正勝[†,‡‡]

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「より高度な認知処理」に基づいた新たな CAPTCHA の導入が強く望まれている．そこで著者らは，「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新しい CAPTCHA を提案する．本稿では特に，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，機械翻訳された文章と人間が作った文章との切り分けを利用した CAPTCHA について検討する．

A Proposal of CAPTCHA using Strangeness in Machine Translation

Takumi YAMAMOTO[†,‡]　J. D. Tygar [††]
Masakatsu NISHIGAKI[†,‡‡]

CAPTCHA is a technique to prevent automatic programs from being able to acquire free Email or online service accounts. However, 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. In this study, to realize a new CAPTCHA, we propose to use a human ability to recognize “strangeness”. As an example, this paper focuses on strangeness in machine translated sentences, and proposes a CAPTCHA which detects malwares by checking if a user can distinguish natural sentences created by human from machine translated sentences.

† 静岡大学創造科学技術大学院
Graduate School of Science and Technology，　Shizuoka University
†† Computer Science Division, University of California, Berkeley
‡ 日本学術振興会特別研究員（DC1）
Research Fellow of the Japan Society for the Promotion of Science (DC1)
‡‡ 科学技術振興機構，CREST
Japan Science Technology and Agency，　CREST

1. はじめに

WEB サービスの発展にともなって，人間と機械を識別するチューリングテストの有用性が益々高まっている．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ためである．チューリングテストは，このような機械（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）を識別するために必須の技術であり，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
Figure 1　An example of a CAPTCHA used for Google Accounts

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（CAPTCHA 解読アルゴリズム，および，PC の CPU パワー）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．

著者らは，人間の「高度な認知処理」として，「違和感」というものに注目した．人間は，自分の経験や常識と少しでも異なるような場面に遭遇すると，「しっくりこない」または「気持ちが悪い」といった感情を，違和感として覚える．より多くの常識を知れば知るほど，また，より豊富な経験を積めば積むほど，違和感を覚える能力は研ぎ澄まされ，非常に些細な違いにも気付くことができるようになると考えられている．すなわち，「違和感を覚える」ことは，高度な認知メカニズムであり，機械による模倣は非常に困難であると期待できる．

そこで，本稿では「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新しい CAPTCHA を提案する．本稿では特に，機械翻訳により生成された文章が有する違和感に着目する．昨今の機械翻訳技術は格段に進歩しているものの，自然な文章を自動的に生成することは現在の最先端技術をもってしても難しい課題であり，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場合がある．これは，機械（機械翻訳プログラム）にとっては，機械翻訳により生成された文章と人間が作った文章との些細な違いを見つけることも困難であるという証拠に他ならない．なぜなら，もし，機械翻訳により生成された文章に対して，人間が覚えるような違和感を機械も認識することが可能ならば，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からである．

以上より本稿では，機械翻訳によって作られた文章の中に存在する違和感を用いた CAPTCHA を検討する．具体的には，機械翻訳により生成された文章（違和感のある不自然な文章）と人間が作った文章（違和感の無い自然な文章）からなる複数の文章群の中から，人間が作った文章のみを選択させることで，ユーザが人間か機械かを判断する．人間であれば，不自然な文章に対して違和感を覚え，自然な文章を容易に選択できる．一方，機械は文章の些細な違和感を理解できないため，不自然な文章と自然な文章とを切り分けられない．

2. 関連研究

文字列の判読能力を試す従来の CAPTCHA においては，既に高機能な 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]．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[5]．本章では，その代表的なものとして Asirra [4]を紹介し，既存の CAPTCHA が抱える課題を示す．

Asirra では，複数の動物の絵（多種多様な背景，角度，ポーズ，照明の違いがある猫や犬の画像）を表示し，その中から特定の動物の絵を選ばせる（図 2）．例えば「猫を選べ」という質問に対し，猫の絵を正しく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，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．

図 2　Asirra [4]の認証画面の例
Figure 2　An example of the authentication window of Assira [4]

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た[5]．この自動プログラムは，「猫の画像によく見受けられる特徴的な情報」や「犬の画像によく見受けられる特徴的な情報」を抽出し，それを機械学習していくことで，猫の画像を高い精度で言い当てることを達成している．

Asirra が破られた原因として，Asirra が，画像の表面的な意味を問う問題形式であったためと考えられる．そこで，本稿では，自然言語の文脈や単語間の関係性といった「内面的な意味」に着目することにより，より高度な CAPTCHA の実現を目指していく．

3. 提案方式

3.1 コンセプト

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．なぜなら，もし，機械翻訳により生成された文章に対して，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からである．すなわち，現在の技術レベルを仮定した場合，機械が機械翻訳により生成された文章と人間が作った文章との些細な違いを見つけることは不可能に近い．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．

そこで本稿では，機械翻訳によって生成された文章に対する違和感を用いた新たなCAPTCHA を提案する．具体的には，機械翻訳により生成された文章（違和感のある不自然な文章）と人間が作った文章（違和感の無い自然な文章）からなる複数の文章群の中から，自然な文章を選択させることで，ユーザが人間か機械かを判断する．人間であれば，文章群の中に含まれている自然な文章を容易に選択できる．一方，機械は自然言語の些細な意味等を理解できないため，不自然な文章と自然な文章とを完全に切り分けられない．本稿では，提案方式を SS-CAPTCHA（CAPTCHA using Strangeness in Sentences）と呼ぶ．SS-CAPTCHA の概観を図 3 に示す．以降では，認証に用いる文章の準備，および，認証手順について説明する．

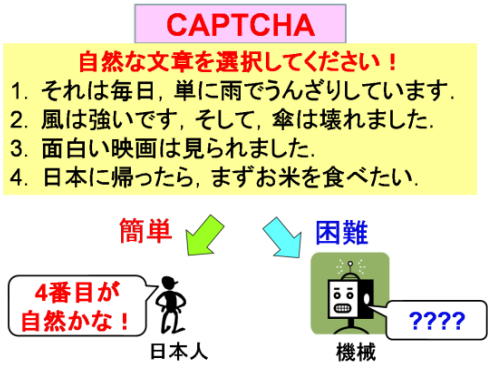

図 3　SS-CAPTCHA の概観
Figure 3　An overview of our SS-CAPTCHA

3.2　文章の準備

SS-CAPTCHA も CAPTCHA である以上，認証に用いる文章（人間が作った「自然な文章」と機械翻訳により生成された「不自然な文章」）を自動的に準備する仕組みが必要である．本節では，「自然な文章」の収集方法，および，「不自然な文章」の生成方法について説明する．以降，本稿では人間が作成した自然な文章，および，機械翻訳により生成された不自然な文章をそれぞれ NS(Natural Sentence)，GS(Garbage Sentence)と呼ぶ．

3.2.1 NS（自然な文章）の収集

インターネット上には人間が作成した文章が無数に存在するが，これらを提案方式における 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（マルウェア）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（WEB 検索によって当該文章が見つかれば NS である）ことが可能だからである．

そこで本稿では，ユーザが SS-CAPTCHA を利用する（以下，「認証を行う」という）たびに，当該ユーザに文章の作成を依頼し，新しい NS を随時蓄えていくという方法を採用する．今回は，ユーザができるだけ容易に文章を作成できるように，文章作成時に画像を 1 枚表示し，その画像を説明する文章を作成してもらうこととした．その際の画像は，画像共有サイトからランダムに抽出される．以降，ユーザが作成した文章を CS（Collected Sentence）と呼ぶ．

なお，当然のことながら，SS-CAPTCHA を利用するユーザが人間であるとは限らない．また，人間であるとしても，そのユーザが自然な文章を正しく入力してくれるとも限らない．よって，認証の際に入力された文章が NS として適正な文章（十分自然な文章）であることを確認する必要がある．この「入力された文章の評価」をパスした CS だけが，NS として SS-CAPTCHA の文章データベースの中に登録されていく．文章の評価方法については 3.3 節で詳しく述べる．

3.2.2 GS（不自然な文章）の生成

不自然な文章 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．

母国語が日本語であるユーザに対する SS-CAPTCHA を例に，提案方式における GS の生成手順を以下に示す．

STEP 1.　システムは初期文章 S_0 を用意する．S_0 は人間が作った自然な文章であり，システムの文章データベースに登録されている NS でも，インターネット上にある文章でも良い．また，S_0 は日本語である必要はないが，その言語の文章を日本語に変換する機械翻訳ツールが利用可能である必要がある．なお，システ

ムは S_0 が何語（L_0）で記述されているかを知っているものとする．

STEP 2. システムは機械翻訳プログラム($MT_{L0->L1}$)を使って，初期文章 S_0 を言語 L_0 から L_1 （$L_0 \neq L_1$）に翻訳する．翻訳語の文章 S_1 は $MT_{L0->L1}$ (S_0)と表現できる．

STEP 3. システムは STEP 2 をランダムな回数(r 回) 繰り返す．繰り返しの度，翻訳語の言語はランダムに決定される．例えば，S_5 は以下のようにいくつかの方法で作成可能である．

$S_5 = MT_{露->中}\ (MT_{韓->露}\ (MT_{仏->韓}\ (MT_{英->仏}\ (MT_{日->英}\ (S_0)))))$

$S_5 = MT_{日->英}\ (MT_{英->日}\ (MT_{日->英}\ (MT_{英->日}\ (MT_{日->英}\ (S_0)))))$

STEP 4. システムは STEP 2 で生成された文章 S_r を日本語に翻訳し，生成された文章 $S_{r+1} = MT_{Lr->日}(S_r)$を GS (不自然な文章)として認証に用いる．

繰り返し回数(r)や組み合わせる言語の数を増やすほど，より不自然な GS が生成されることが予想される．

3.3 SS-CATCHA の仕組み

本稿で提案する SS-CAPTCHA は，「収集」および「認証」の 2 つのフェーズからなっている．「収集」フェーズは新たに NS（自然な文章）を追加していくために実行される．一方，「認証」フェーズには 2 つの目的がある．1 つ目は「当該ユーザが人間なのか機械なのかを切り分けること」，すなわち本来の CAPTCHA の機能である．2 つ目は「他のユーザの収集フェーズにて入力された文章が NS（自然な文章）としてふさわしいかを評価すること」である．以下に，SS-CAPTCHA の収集フェーズと認証フェーズの仕組みを示す．またそれらの仕組みを図に表したものも図 4〜図 6 に示す．収集フェーズと認証フェーズはセットになっており，ユーザは収集フェーズを行った後，認証フェーズを行う．

●収集フェーズ

STEP 1. システムはユーザに一枚の画像を提示する．この画像は画像共有サイトからランダムに検索される[a]．ユーザは与えられた画像を元に，それを説明する文章を作成し，システムに入力する．今回は，短く単純な文章を作成してもらうこととした．

STEP 2. システムは，STEP 1 にて入力された文章を WEB 検索エンジンで検索する（完全一致検索）．同じ文章が WEB 上に見つかった場合は，入力された文章を破棄する．必要によっては，ユーザに文章の再入力を求めるようにしてもよい，入力された文章を WEB 検索することによって，入力された文章が新たに生成された文章であることを確認している．

STEP 3. STEP 2 を通過した文章は CS（Collected Sentence）としてシステムの文章データベースに蓄えられ，当該ユーザとは異なるユーザが実行する認証フェーズにて，文章の質（自然な文章であるか否か）が評価される．

●認証フェーズ

STEP 1. システムはユーザに「NS：自然な文章」，「GS：機械翻訳により生成した不自然な文章」，「CS：当該ユーザとは別のユーザの収集フェーズにて入力された文章」とからなる文章群を表示する．表示する NS の数を N, GS の数を G, CS の数を C と表す．文章群に含まれる文章の総数を S とする（S=N+G+C）．

STEP 2. ユーザは文章群の中から違和感の無い文章を M 個選択する．

STEP 3. 文章群の中に含まれる N 個の NS の中から θ_{NS} 以上の NS を選択することができれば，当該ユーザは「人間」として判定され，そうでなければ，「機械」として判定される．

STEP 4. STEP 3 にて，人間であると判定されたユーザが選択した CS は，自然な文章である可能性が高いため，当該 CS の評価値 V を 1 点増やす．人間であると判定されたユーザが選択しなかった CS は，自然な文章でない可能性が高いため当該 CS の評価値 V を 1 点減らす．STEP 2 にて機械であると判定されたユーザの選択は CS の評価値 V に影響を与えない．また STEP 4 では，CS だけでなく GS に対しても，CS 同様に文章の質の評価を行う．これは，文章によっては機械翻訳により生成された文章が自然な文章になることもあり，そのような GS が継続してそれ以降の認証の中で使われることを防ぐためである．なお，CS および GS の評価値 V の初期値は 0 である．

CS は，K 人のユーザによって評価される．すなわち，あるユーザの収集フェーズにて入力された文章（CS）は，その後，任意のタイミングで計 K 人のユーザの認証フェーズの中で表示される．K 人のユーザによる評価の後，CS の評価値 V が θ_{CS} 以上であれば，CS は「自然な文章 NS としてふさわしい」と判断され，以降は NS としてシステムの文章データベースに登録される．そうでなければ，CS は破棄される（または GS として利用する）．同様にして GS も評価され，GS の評価値 V が θ_{GS} 以上であれば，GS は「不自然な文章 GS としてふさわしくない」と判断され，GS として利用されない（または NS として用いられる）．

a 不適切な画像や言葉での説明が困難な抽象的な画像が抽出されないよう，フィルタリング機能やタグ検索機能を有する画像共有サイトを利用すると良いだろう．

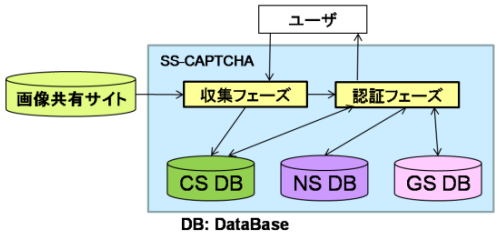

図 4　SS-CAPTCHA の仕組み
Figure 4　The scheme of SS-CAPTCHA.

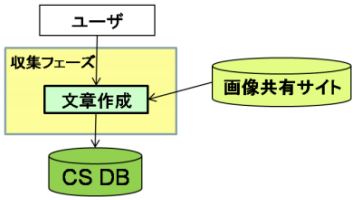

図 5　収集フェーズの仕組み
Figure 5　The scheme of collecting phase SS-CAPTCHA.

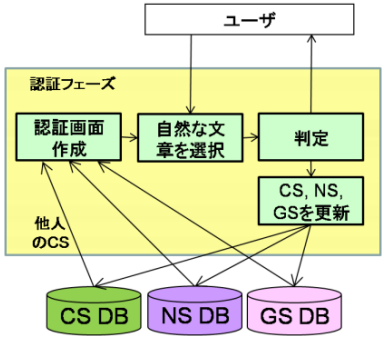

図 6　認証フェーズの仕組み
Figure 6　The scheme of authentication phase in SS-CAPTCHA

4.　検証実験

SS-CAPTCHA の実現可能性を確かめるために，基礎実験を行い評価する．本実験の被験者は本学情報学部学生 8 名である．

4.1　自然な文章の収集

●実験の目的

SS-CAPTCHA の収集フェーズにて画像を元に自然な文章を作成することが，ユーザ（人間）にどの程度負荷を与えるのかを検証する．

●実験方法

本実験で用いた画像は画像共有サイト flickr [6]を使ってあらかじめ収集したものを用いた．画像を収集する際に，「culture」，「food」，「sports」など一般的なキーワードを元にタグ検索を行った．キーワードを用いた理由は，ユーザがある程度意味をつけ易いと考えられる画像を抽出するためである．

被験者には画像が 1 枚ずつランダムに与えられ，画像を元に適当な長さの文章を作成してもらった．なお，できるだけ「話し言葉」ではなく「書き言葉」で文章を作成するよう被験者に指示した．

●実験結果

被験者 8 名から計 161 個の文章が得られた[b]．1 文章当たりの作成時間の平均は 45.11 秒であった．

また，本実験で得られた文章を機械翻訳にかけた場合に，SS-CAPTCHA にとって望ましい GS（不自然な文章）が生成されるかを確かめた．その結果を表 1 に示す．機械翻訳による GS の生成方法は次の 3 種類である．本実験で用いた機械翻訳ツールはエキサイト 翻訳[7]を利用した．

①　日本語→英語→日本語
②　日本語→イタリア語→日本語
③　日本語→英語→イタリア語→日本語，日本語→イタリア語→英語→日本語

ここで，本実験で英語およびイタリア語を用いた理由は，事前調査にて英語，イタリア語の順で翻訳精度が高いという直感的感触を得たためである．なお，手法①，②では 1 つの文章から 1 つの GS が生成されるが，手法③では 1 つの文章から 2 つの GS が生成される．すなわち，161 個の日本語文章から生成された GS の総数は，手法①，②ではそれぞれ 161 個，手法③では 322 個である．

表中，「A」は，翻訳後の日本語文章のうち，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翻訳後の文中に含まれていた文章の数を表す．これは翻訳プロ

b 各被験者あたり 20 個の文章を作成してもらったが，手違いで 1 名の被験者のみ 21 個の文章となった．

グラムで用いられる辞書に存在しない単語（または表現）が入力されたため，その部分がローマ字で表記されたり，記号で置き換えられたりしたものと考えられる．また表中，「B」は，「A」に該当する文章を除いた全 GS の中で，人間が作った文章と同程度に違和感の無い文章（自然に見える文章）の数である．今回は，違和感があるか無いかの判断は著者自身が行った．その際に，「文の意味」，「文の流れ」，「単語の修飾のされ方」などに注意して判断した．

表 1　GS の質について
Table 1　Quality of Garbage Sentence created by processing technique ① ~ ③

①		②		③	
A	B	A	B	A	B
16	15	23	11	54	14

●考察

文章作成には平均でも 45.11 秒も要してしまうことから，ユーザの負担は大きいことが見て取れる．被験者からの聞き取り調査では，「言葉で説明することが難しい画像がある」という意見が多数あり，文章を作り易い適切な画像を用意する必要がある．

また表 1 より，手法①～③によって生成された GS のうちの約 2 割の文章が A または B に分類されていることが分かる．A に分類される文章は，不自然な文章であるということを機械的な判定することが容易であると思われるため，GS から除外すべきであると考えられる．一方，B に分類される文章は，NS（自然な文章）と間違えられるため，SS-CAPTCHA の認証精度を劣化させる原因になる．よって，3.3 節で説明したように，認証フェーズにおいては GS に対してもその不自然らしさを評価していく必要があることがわかる．

4.2 認証実験

●実験の目的

機械翻訳により生成された文章と人間が作った文章の切り分けが人間にとって容易かどうかを確認する．

●実験方法

本実験では，4.1 節で被験者に作成してもらった文章を NS（自然な文章），その文章から手法①,②,③の機械翻訳によって得られた文章を GS（不自然な文章）として用いる．ここで，4.1 節の手法①,②,③により得られた GS の集合をそれぞれ ***GSs(1)***,***GSs(2)***, ***GSs(3)***，4.1 節で被験者から得られた文章の集合を ***NSs*** と示す．本実験では GS 生成手法ごと別々に認証実験を行った．手法①を例に実験手順を以下に示す．

STEP 1. システムは ***NSs*** の中からランダムに 5 個の NS を選択する（N=5）．その際，当該被験者が作成した NS は選択しない．
STEP 2. システムは ***GSs(1)***の中からランダムに 10 個の GS を選択する（G=10）．その際，当該被験者が作成した NS から生成された GS，STEP 1 で既に選択されている NS から生成された GS，ならびに，4.1 節で「A」に分類された GS は選択しない．（手法③の場合は，1 つの NS から 2 つの GS が生成されるが，1 つの NS から作成された GS が 2 つとも選択されないようにした．）
STEP 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（図 7）．
STEP 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（M=5）．

被験者には STEP1~STEP4 を各 GS 生成手法につき計 10 回実行してもらう．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 30 回実行する．1 回の選択作業の度に GS 生成手法がランダムに選択される．

●実験結果

実験結果を表 2 に示す．表中，「選択時間」は，被験者が 15 個の文章の中から違和感の無い 5 個の文章を選択するまでに要した時間の平均である．「成功率」は，全実験のうち，被験者によって選択された 5 個の文章の中に NS が θ_{NS} 個以上含まれていた場合の割合である．また，「C」は，4.1 節にて「自然な文章」として評価された GS（以下，「自然な GS」と呼ぶ）をそのまま用いて認証を行った場合の実験結果である．一方，「D」は，自然な GS は 3.3 節に示した仕組みによって NS として利用されることを想定し，「自然な GS」も NS に含めた上で認証成功率を算出した結果である[c]．D は，認証成功率を再評価しただけで，実験そのものをやり直しているわけではないことに注意されたい．

●考察

イタリア語のみを用いた GS 生成手法②の方が，英語のみを用いた GS 生成手法①よりも認証成功率が高い．また複数の言語（英語とイタリア語）を用いた GS 生成手法③の方が，英語のみを用いた GS 生成手法①よりも認証成功率が高い．これは，翻訳精度が（英語よりも）悪いイタリア語を使ったことによる効果だと考えられる．

今回の実験では，「自然な GS」を排除していないため，被験者は違和感のない文章を選択するにあたって迷いが生じていたと推測される．被験者が文章の選択に平均 60 秒強も要してしまっており，また，（「自然な GS」を NS に含めない場合の）θ_{NS}=5 に対する認証成功率が 5 割未満に留まってしまっているのは，この「ユーザの迷い」

c 選択された 5 個の文章の中に，NS と「自然な GS」が合わせて θ_{NS} 個以上含まれていた場合に，認証成功とする．

が原因となっていると考えられる．「自然な GS」を NS に含めると，θ_{NS}=5 の場合の認証成功率は 8 割程度に向上するが，依然として十分な認証精度とは言えない．これは，「不自然な NS」の存在が原因となっている．以上より，3.3 節の仕組みによって「自然な GS」や「不自然な NS」を正しく評価し，適切に取り扱うことの重要性が確認された．

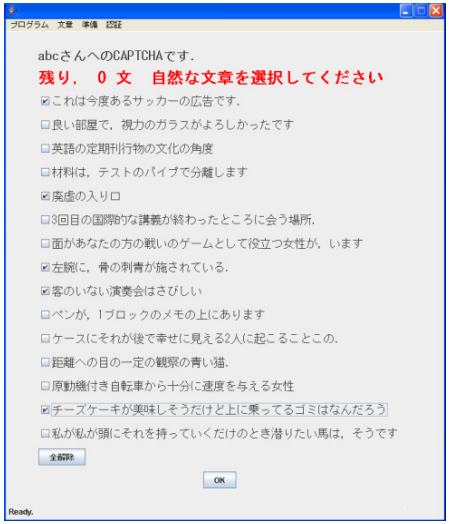

図 7　認証画面の例

Figure 7　An example of an authentication window of SS-CAPTCHA

表 2　認証成功率と選択時間

Table 2　Authentication success rate and time required for sentence selection

		GS 生成手法					
		①		②		③	
		C	D	C	D	C	D
θ_{NS}	5	17.50%	52.50%	46.25%	71.25%	28.75%	71.25%
	4	76.25%	93.75%	91.25%	98.75%	92.50%	98.75%
	3	93.75%	100.00%	100.00%	100.00%	100.00%	100.00%
	2	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
選択時間		69.36 [sec]		66.73 [sec]		60.37 [sec]	

5.　検討

本章では，提案方式の利便性および安全性について，4 章の基礎実験で得られた知見をもとに議論する．

5.1　利便性について

4 章の実験では，文章作成に平均約 45 秒を，認証に平均約 60 秒を要するという結果であった．すなわち，SS-CAPTCHA を 1 回実行すると 2 分弱もの時間がかかってしまう．利便性向上のためにも，「収集フェーズ」および「認証フェーズ」における工夫が必要である．

「収集フェーズ」においては，「言葉で説明することが難しい画像がある」という被験者からの多数の意見があったことからも，文章作成用の画像をどのように選定するかが重要な検討事項となってくる．また，ユーザに文章を作成してもらうことなく，自然な文章を自動的に収集する方法の検討も今後必要になってくると考えられる．

「認証フェーズ」においては，4.2 節で述べたとおり，違和感のない文章の選択に迷いが生じないように，適切な文章群を提示することが必須である．これは，3.3 節の仕組みに基づいて「自然な GS」や「不自然な NS」を排除することで，大幅な改善が図られると予想される．また，「1 回の認証画面に表示される文章が 15 個と多いため，絞り込みに悩まされた」という被験者の意見もあった．1 回の選択における文章の数をどこまで減らしても SS-CAPTCHA が正しく機能するかについても検討していく必要があるだろう．

また，鈴木ら[9]が提案しているように，正規ユーザ（人間）にとって「心地良い（エンターテイメント性を有している）」CAPTCHA であれば，多少時間を要する方式であったとしても，煩わしさを感じさせない CAPTCHA を実現することができるかもしれない．これらの改善が適応可能か今後の課題として議論していく予定である．

5.2　安全性について

5.2.1 ランダム選択による攻撃

4.2 節で用いた実験システム（15 個の文章の中から 5 個の文章を選ぶ）において，ランダム選択により正しい文章（5 個の NS）を選択する確率は，$1/{}_{15}C_5$（1/3003）である．しかし，4.2 節の実験結果からもわかるとおり，人間であっても 5 個の NS を全て正しく選択できるわけではない．θ_{NS}=4 にすることで人間の受入率を高くすることができるが，その結果，ランダム選択により 4 個の NS を選択する確率は ${}_5C_4 / {}_{15}C_4$（1/273）となり，総当たり数は減少してしまう．利便性と安全性の双方を考慮した適切なパラメータ設定を検討していく必要がある．

5.2.2 機械による攻撃

SS-CAPTCHA は，「機械翻訳プログラムは飛躍的な発展を遂げてきたものの，依然として自然な文章を自動的に作ることが困難であるという事実から，機械は機械翻訳を用いて作成された文章と人間が作った文章との些細な違いを見つけることも困難なのではないか」という仮定に基づいている．しかし，「違和感の無い自然な文章を自動的に作ること」と「違和感のある不自然な文章を見つける」ことは，本当に機械にとって同程度に困難な課題であるかには疑問が残る．これについては今後早急に検討し，SS-CAPTCHA が機械にとって本当に解読困難な問題であるかについても調査していく予定である．

5.2.3 人間の能力を活用した CAPTCHA の解読

人間の能力を活用した CAPTCHA の解読が，大きな社会問題となってきている．この問題は大きく分けて「低賃金労働力による CAPTCHA の解読[9]」および「ポルノサイト閲覧者による CAPTCHA の解読[1，10]」に分けられる．

●低賃金労働力による CAPTCHA の解読[9]

非常に安い賃金で労働者を雇い，解読したい大量の CAPTCHA を，低賃金労働者に送り，彼らに回答してもらうことで，CAPTCHA を無効にしようとする攻撃である．

本攻撃に対しては，提案方式はある程度の耐性を有していると考えられる．なぜなら，例えば日本語を用いた CAPTCHA の場合，一般に，日本語を母国語とするユーザでなければ，日本語の文章の些細な意味の違いに気付くことは難しいからである．言い換えると，「アラビア数字やアルファベットを読むことができればよい」という既存の CAPTCHA 解読に必要な労働者の雇用条件に比べ, SS-CAPTCHA の解読に必要な労働者の雇用条件が格段に厳しくなる．

●ポルノサイト閲覧者による CAPTCHA の解読[1]

ポルノサイト訪問者にポルノ画像を見せる代わりに，解読したい CAPTCHA に回答してもらうことで，CAPTCHA を無効にしようとする攻撃である．

本攻撃は提案方式においても大きな問題であり，解決策は未だ見つかっていない．今後，ポルノ以外のサービスサイトを利用して同様の CAPTCHA 解読を試みる不正者が増えてくると推測されることから，この種の攻撃に対して早急に検討していかなければならない．

6. おわりに

本稿では「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新しい CAPTCHA のコンセプトを提案した．違和感の一例として，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，機械翻訳された文章と人間が作った文章との切り分けを利用した SS-CAPTCHA について検討した．

今後は，本稿で詳しく検討することができなかった「自然な GS」および「不自然な NS」を排除する仕組みの評価，GS の生成手法（他の言語との組み合わせや，機械翻訳の多重化における GS の評価）についての検討，ならびに，提案方式の安全性についての評価を実施していく予定である．

謝辞 本研究は科研費（No.20-6290）の研究助成を受けている．また，本研究は一部，（財）セコム科学技術振興財団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 Naive 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) Welcome to Flickr - Photo Sharing, http://www.flickr.com/
7) エキサイト 翻訳, http://www.excite.co.jp/world/
8) 鈴木徳一郎，山本匠，西垣正勝：4 コマ漫画 CAPTCHA の提案，2009 年暗号と情報セキュリティシンポジウム予稿集，CD-ROM（論文 No.3D3-3），2009.
9) Google の CAPTCHA 実験が的外れな理由，ZDNet Japan, http://japan.zdnet.com/sp/feature/07zeroday/story/0,3800083088,20392346,00.htm?ref=rss
10) 今度はポルノ画像をエサに--スパマー対フリーメールサービスの「イタチごっこ」，CNET Japan，http://japan.cnet.com/news/sec/story/0,2000056024,20065869,00.htm