人間ロボット判別テストのバリアフリー化のための言語的作問とその自然文生成技法

山口 通智 †　　岡本 健 ‡

筑波技術大学大学院技術科学研究科
305-8521　茨城県つくば市春日 4-12-7
† ym123202@cc.k.tsukuba-tech.ac.jp, ‡ ken@cs.k.tsukuba-tech.ac.jp

あらまし 現在、実社会で用いられている CAPTCHA は、不明瞭な画像や音声を解釈をさせる手法を用いているため、知覚障碍者の利用には障壁がある。この障壁を取り除くため、言語的作問の利用が提案されており、機械合成文と自然文の判別はその一例である。数に制限のない作問のためには、公開文を含めた多量の素材文章が必要になるが、攻撃者が検索により自然文を検出する問題がある。本研究では、複数階数のマルコフ連鎖に基づき新規に生成された機械合成文を自然文として用いることで、問題の解決を図る。さらに、実験を通して、その有効性を確認した。

An Advanced CAPTCHA using Generation of Accessible Verbal Questions

Michitomo Yamaguchi†　　Takeshi Okamoto‡

Department of Technology and Sciences, Tsukuba University of Technology,
Kasuga 4-12-7, Tsukuba City, Ibaraki, JAPAN, #305-8521.
† ym123202@cc.k.tsukuba-tech.ac.jp, ‡ ken@cs.k.tsukuba-tech.ac.jp

Abstract The current CAPTCHA mainly uses perceptual recognition with images or sounds, but it is difficult for visually or hearing impaired people to leverage it. In this paper, we propose a new identification scheme. Our scheme uses a verbal question which is identification of natural sentences and word salads. We generate natural sentences as word salads which are generated by the Markov chain with multi-state. Such sentences are generated almost unbounded volume by using online documents and it is hard for adversaries to find them by search engine. We also evaluate our scheme by some experiments.

1　はじめに

1.1　研究の背景

2 者間の対話において、回答者が人間であるか、あるいは人工知能による自動プログラム（通例「ロボット」と称される）であるかを判別するテストは「人間ロボット判別テスト」と呼ばれている。このテストは、今日のネット上において個人認証手続きで多用されている。その目的は、パスワード認証に対する総当たり攻撃や、アカウントの大量生成、多重投票といった不正を目的とするロボットが、短時間に大量にアクセスできないようにするためである。

人間ロボット判別テストでは一般に、人間には容易に解けるが、現在のロボットでは解くことが難しい AI (Artificial Intelligence、人工知能) 問題を用いる。このようなテストの代表例として、歪んだ文字画像を読み取らせる CAPTCHA (Com-

図 1: 一般的な CAPTCHA の例

pletely Automated Public Turing Test To Tell Computers and Humans Apart) [1, 8] がある。図 1 のように、歪曲やノイズが付加された文字列の画像を作り、利用者にその文字列を判読させ回答させる。これは、人間の画像認識能力が現状のロボットより上回っているという仮定に基づく。また、視覚障碍者向けの代替として、変形した音声を利用者に聞き取らせる方式もある。

本稿で取り組む問題は、既存の人間ロボット判別テストの多くが、特定の知覚チャンネルに限定して作られているため、その知覚に障碍がある利用者の障壁になっている点である。ここに、知覚チャンネル依存という問題点を解消し、知覚障碍によらず誰でも利用できる人間ロボット判別テストを作成するための要件を示す。

バリアフリー要件： 特定の知覚のみの使用に限定されないこと。

識別性要件： 人間には容易に解けるが、現状のロボットには解答が難しい問題を生成できること。

問題新規性要件： 未使用で新規な問題を無数かつ自動で作れること。

知識非依存性要件： テストの難易度が特定の知識の有無に強く依存しないこと。

1.2 知覚チャンネル非限定テストの関連研究とその問題点

まず、バリアフリー化に関しては、画像型と音声型の CAPTCHA を併用することが考えられる。しかし、音声型 CAPTCHA は、人間には難しい [2, 3] ことが知られいる。人間の正答率は、2009 年の時点でさえ 50% に満たず、近年ではさらに難易度が上昇している。

次に、画像型 CAPTCHA テストを視覚以外に拡張するというアイデアがある。Holman ら [5] は、サイレンや鳥などの身近な事物を画像と音声の両方で提示し、そのいずれによっても回答可能とする方式を提案した。ただし、事物の画像音声の用意の手間に鑑みると、問題新規性要件を十分に満たすことは難しいと考えられる。

問題の難しさの根拠が、特定知覚に限定されない AI 問題を利用するアイディアもある。たとえば、文意文脈の理解能力を試す問題（「文意文脈解釈問題」と称す）は、人工知能や認知科学の分野で古くから研究課題になっており、参考とすべきテストのアイデアが見受けられる。

認知科学における「サリーとアン課題」 [9] では、自分の知識の範囲と他者のそれとを区別できるかを問うテストがある。ディックの SF 小説 [4] では、「このカバンは官給品なんだ。赤ん坊の皮でできている。」などと、異様な内容の文章を聞かせ、異様部分に対する身体的・感情的反応の時間遅れを計測する 「Voigt-Kampff Test」 のアイデアが示されている。文意・文脈の理解と常識の発揮の能力差を利用する例としては先駆的なものと言える。これらはロボットには難しいテストであり、識別性要件を満たす。しかし、自動の作問が難しく、問題新規性要件を満たせるかに疑問がある。

文意文脈解釈問題を用いる CAPTCHA の研究では、人間とロボットの間に、文意や文脈の解釈能力の差があることを利用する。山本ら [10] は、人間が作った文章と機械翻訳により生成される文章との間で、人間が感じる違和感を人間ロボット判別に用いた。同様に、上原ら [11] の、人間が作った文章とマルコフ連鎖により自動合成された文章の識別の研究がある。Christopher [7] は、複数の文の中から内容が関連するものとしないものを選択させる方式を提案した。

この種の作問では、自然な文として、人間が作った文が必要となる。問題新規性要件を満たすためには、その素材となる文章の量は、十分に大きくなければならない。文章量の確保には、インターネットなどで公開された文章の利用が有効である。反面、それは公開情報であるため、攻撃者も同じ文章を検索によって収集できると

いう欠点がある。これでは識別性要件を満たさない。

先行研究では、作問者のみが知りうる秘匿された文章を用いる方式 [11] や、テストの途中で回答者に入力させた文章を用いる方式 [10] がある。これらの方式は、公開文章の利用に比べ、問題新規性を十分に満たすことは難しいと考えられる。しかも、問題新規性を満たす十分な文章量がなければ、これらの方式も識別性要件を満たさない。なぜなら、秘匿された文章であっても、一度問題として使えば、そこで公開されてしまうので、秘匿が成り立たない。問題新規性が十分で無ければ、攻撃者がテストシステムへの出題要求を大量に繰り返すことで、いつかは既知で解答を分析済みの問題文に遭遇できてしまう。後者に関しては、攻撃者も文章入力に参加することで、攻撃者に都合の良い素材文章を登録できてしまう。

1.3 研究の目的

本研究では、節 1.1 に示した 4 つの要件を満たす人間ロボット判別テストの構成を目的とする。

本稿では、バリアフリー要件を満たすため、文意文脈解釈問題を用いる。さらに、知識非依存性要件を満たすため、人間が作った文と機械が生成した文との間で人間が感じる違和感を利用した方式 [10, 11]（「不自然な文の識別問題」と称す）に着目する。しかしながら、前述の通り、不自然な文の識別問題は、識別性要件と問題新規性要件の達成が課題となる。

本稿では、この問題を解決するため、階数 N の大きいマルコフ連鎖から生成される機械合成文を、自然な文として扱うことを検討する。まず、節 3 にて単純な方式の問題点を指摘する。次に、節 4 にて提案方式を、節 5 にて実験用プログラムによる評価を示す。

2 準備

2.1 表記

本稿で用いる記法について示す。$\mathcal{S}$ を集合とする。$s \overset{D}{\leftarrow} \mathcal{S}$ は、$\mathcal{S}$ から確率分布 D で要素 s を選択することを表す。特に $D = \$$ の場合は、一様ランダムな分布から要素 s を選択することを表す。a, b を変数、リテラル、数字のいずれかとし、c を変数とする。$c \leftarrow a \circ b$ は、a, b に対する演算 $\circ$ の結果を c に代入することを表す。$c \leftarrow a$ は、a を c に代入することを表す。$|A|$ は、A が集合や配列であればそのサイズを、A が文字列ならばその文字数を表す。$[a, b]$ は、$a \leq x$ かつ $x \leq b$ となる整数 x の範囲を表す。

2.2 N 階マルコフ連鎖

N 階マルコフ連鎖は、次式に示されるような、直前 N 個の状態に依存して次の状態が決定される確率過程である。

$$
\begin{aligned}
& \Pr[\mathrm{X}_{n+1} = x | \mathrm{X}_n = x_n, \ldots, \mathrm{X}_0 = x_0] \\
= \; & \Pr[\mathrm{X}_{n+1} = x | \mathrm{X}_n = x_n, \ldots, \mathrm{X}_{n-N+1} = x_{n-N+1}]
\end{aligned}
$$

2.3 ワードサラダ

ワードサラダとは、機械合成文の一種である。本稿で取り扱うワードサラダは、節 2.2 に示したマルコフ連鎖を用いた言語モデル（「マルコフ連鎖モデル」と称す）により生成される文を対象とする。

マルコフ連鎖モデルを生成するには、ある文書集合（「文章源」と称す）から取得した文章（「素材文章」と称す）を入力として用いる。モデルの生成プログラムは、素材文章に形態素解析を施し、ある N-gram の形態素とそこから連鎖する形態素の組み合わせを取得し、その頻度情報と合わせてモデルに登録する。プログラムは、文中のある形態素が直前の N-gram の形態素により決まる連鎖型共起表現であると仮定し、マルコフ連鎖モデルから連鎖する形態素の取り出しを再帰的におこなうことで、ワードサラダを生成する。

本稿では、N 階マルコフ連鎖モデルから生成されたワードサラダを、「N 階ワードサラダ」と表す。また、生成に複数階数のマルコフ連鎖モデルを使用した場合は、「$[N_L, N_H]$ 階ワードサラダ」のように、階数の範囲 $[N_L, N_H]$ を用いて表す。

表 1: 形態素 N-gram のパタン数（左）と N 階マルコフ連鎖の平均パタン数（右）

素材文章	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$
A	184, 2.0	372, 1.1	422, 1.0†	435, 1.0†	435, 1.0
B	4361, 4.7	19656, 1.8	35204, 1.3	43055, 1.1	45524, 1.0†
C	10961, 4.4	47688, 1.8	83785, 1.2	99911, 1.1	104308, 1.0†
D	6276, 4.3	27065, 1.8	47665, 1.2	58439, 1.1	62981, 1.0†
E	1308, 2.8	3603, 1.4	4828, 1.1	5256, 1.0†	5363, 1.0†

†：小数点第 2 位は 0 ではない

3　単純な方式の問題点

N 階ワードサラダは、素材文章から抜き出された N-gram の形態素で構成される。従って、階数 N が大きければ、小さい場合に比べて、人間には自然な文と認識されやすい。しかしながら、ワードサラダを自然な文として利用するには、単に N を大きくするだけでは解決しない。

表 1 に、ある素材文章の形態素 N-gram のパタン数と、N-gram の形態素から連鎖する形態素の平均パタン数を示す。表 1 の素材文章は、青空文庫を文章源とし、そこからランダムに収集された 5 種である。$N = 5$ における連鎖する形態素の平均パタン数は、ほぼ 1.0 である。つまり、これらの素材文章から 5 階ワードサラダを生成すると、単に素材文章の一部が切り出される確率が高い。攻撃者は、検索結果から問題に回答できてしまう。

対策としては、階数を小さくしていくことで、連鎖する形態素の平均パタン数を増やし、素材文章に存在しない文字列を、新規文（「新規ワードサラダ」と称す）として生成できる。しかし度が過ぎると、人間は、不自然な文と識別できなくなる。また、連鎖する形態素の平均パタン数は、$N = 3, 4$ でも 1.0 近くに収束するので、適切な階数調整ができない可能性がある。

階数による文の自然さの調整は、別な問題もある。例えば、1 階ワードサラダを不自然な文とし、4 階ワードサラダを自然な文とする。これらのワードサラダをウェブ検索すると、使用した文章源が検索結果に現れる頻度は、後者の方が高いと推測される。検索結果に違いがあるならば、攻撃者はそれを手がかりにできる。

4　提案方式

4.1　モデルの生成方法

本研究では、複数階数のマルコフ連鎖モデルを用いて、人間には比較的自然な文と感じる新規ワードサラダの生成をおこなう。これは、小さい階数で生成される文の新規性と、大きい階数で生成される文の自然さを併せ持つことを狙う。さらに、マルコフ連鎖モデルを、品詞・活用形の種類にも拡張する。これは、新規ワードサラダの生成パタンを増やし、不自然な文との検索結果の相違を無くすために導入する。これについては、節 4.2 で詳細を述べる。

モデルの構成方法を示す。形態素によるマルコフ連鎖モデルを C_0、品詞・活用形によるものを C_1 とする。これは、ハッシュ鍵 *key* に対して複数の値をもつハッシュ構造として構成する。C_0 であれば、ハッシュ鍵は N-gram の形態素であり、ハッシュ値はそれから連鎖して文章に登場する形態素とその頻度情報である。C_1 は、形態素の代わりに、その品詞・活用形の情報を使用する。品詞・活用形の情報は、形態素解析 [6] で得られたものを利用する。その際、形態素とその品詞・活用情報の対応マップ C_2 を合わせて構成する。C_2 も、ハッシュ鍵に対して複数のハッシュ値をもつ構造となる。

文章源を $\mathcal{M}$ とし，そこから取得した素材文章を段落[1]ごとに分割したものを m する。形態素解析により m の形態素配列 *mor* を取得する。*mor* の i 番目の形態素を *mor*[i] とする。*mor* のサイズを +1 し、追加要素として、終端を表す

[1]実装したプログラムでは、MS-WORD のパラグラフに代表されるような、改行区切りとした。

特殊記号（終端記号と称す）を配列の最後に格納する。階数の最大値を N_H とし、$N \in [1, N_H]$ に対して、次の処理で C_0 を構成する。

(1) $i \leftarrow 0$ とする。

(2) $i + N < |mor|$ ならば (3) に進む。そうでなければ、処理を終了する。

(3) $key \leftarrow (mor[i], \ldots, mor[i+N-1])$ とする。$mor[i+N]$ とその頻度情報に対して、次の判定をおこなう。

- key が C_0 に未登録であれば、key をハッシュ鍵、形態素 $mor[i+N]$ と頻度情報 1 をハッシュ値として登録する。
- key が C_0 に登録済みだが、形態素 $mor[i+N]$ がハッシュ値として未登録であれば、それと頻度情報 1 をハッシュ値として追加する。
- key と $mor[i+N]$ の組みが登録済みであれば、対応する頻度情報を +1 する。

(4) ハッシュ鍵として登録された最初の形態素 $mor[i]$ が自立語の場合は、文頭可能属性としてマークを付ける。

(5) $i \leftarrow i + 1$ とし、(2) に進む。

C_1 の構成は、C_0 と同様になるが、形態素の代わりにその品詞・活用形を利用し、(4) の処理を省略する。

4.2 ワードサラダの生成方法

まず、階数 N を固定したワードサラダの生成法を示す。$val \overset{D}{\leftarrow} C(key)$ は、key に対するハッシュ値として登録された要素を、その頻度分布 D に従い C からランダムに取得することを表す。例えば、key に対して val_0 が頻度 7、val_1 が頻度 3 と登録されていた場合、val には 70% の確率で val_0、30%の確率で val_1 が代入される。ary を、1-gram 形態素を要素とする配列とする。

(1) 文頭要素の取り出し：C_0 からランダムに文頭可能属性を持つハッシュ鍵 key_0 を取り出す。key_0 は N-gram の形態素であることに注意を要する。$val_0 \overset{D}{\leftarrow} C_0(key_0)$、$ary \leftarrow (key_0, val_0)$ とする。

(2) 終了判定：val_0 が終端記号であるか、ary に格納された総文字数がしきい値 ℓ_{min} 以上ならば処理を終了し、ary に格納された文字列を返す。そうでなければ、(3) に進む。

(3) 連鎖する形態素の取得：ary の最後から N 個の要素をハッシュ鍵　key_0 として取り出す。$val_0 \overset{D}{\leftarrow} C_0(key_0)$、$ary \leftarrow (ary, val_0)$ とし、(2) に進む。

次に、複数階数 $[N_L, N_H]$ への対応を述べる。概要としては、階数 N_L, N_H を交互に適用してワードサラダを生成する。しかし $N_H - N_L > 1$ だと、N_H-gram の形態素から連鎖するパタンが C_0 に存在しない可能性がある。その場合には、階数を減らしていけば連鎖パタンが見つかるが、N_H がほとんど適用できず、不自然な文しか生成できない。かといって N_H, N_L の差が少なければ、人間に自然と感ずる新規ワードサラダの生成がしにくい。よって、本方式では、形態素での連鎖パタンがない場合には、品詞・活用形が同種となる連鎖パタンを探索する。

(1a) 文頭要素の取り出し：C_0 からランダムに文頭可能属性を持つハッシュ鍵 key_0 を取り出す。$val_0 \overset{D}{\leftarrow} C_0(key_0)$、$ary \leftarrow (key_0, val_0)$ とする。

(1b) 初期階数の設定：key_0 は $n \in [N_L, N_H]$ となる n-gram の形態素である。$N \leftarrow n$ とする。

(2a) 終了判定：val_0 が終端記号であるか、ary に格納された総文字数がしきい値 ℓ_{min} 以上ならば処理を終了し、ary に格納された文字列を返す。そうでなければ、(2b) に進む。

(2b) 階数の変更：$N \neq N_H$ であれば $N \leftarrow N_H$ とし、そうでなければ $N \leftarrow N_L$ とする。

(3a) 形態素による連鎖の可否判定：ary の最後から N 個の要素をハッシュ鍵　key_0 として取り出す。key_0 が C_0 に登録されていれば (3b) に、そうでなければ (3c) に進む。

(3b) 連鎖する形態素の取得：$val_0 \overset{D}{\leftarrow} C_0(\,key_0)$、$ary \leftarrow (key, val_0)$ とし、(2a) に進む。

(3c) 品詞・活用形による連鎖の可否判定：$key_1 \leftarrow C_2(\,key_0)$ とし、N-gram 形態素 key_0 に対応する品詞・活用形の組み合わせ key_1 を得る。C_1 に対して、ハッシュ鍵 key_1 の照合をし、登録されていれば (3d) に、そうでなければ $N \leftarrow N - 1$ とし (3a) に進む。

(3d) 品詞・活用形が同種となる形態素の取得：$val_1 \overset{D}{\leftarrow} C_1(\,key_1)$、$val_0 \overset{D}{\leftarrow} C_2(\,val_1)$、$ary \leftarrow (\,ary, val_0)$ とし、(2a) に進む。

品詞・活用形の利用は、文章源の検索に対しても有効であると考える。このように生成されたワードサラダは、文法的には形態素のみの生成法と相違はないが、素材文章に存在しない形態素の組を含む。よって、検索により文章源が検出される頻度が下がり、検索結果の分布が 1 階ワードサラダなどに近づくと期待できる。

なお、品詞・活用形のマルコフ連鎖モデルのみを用いたワードサラダを採用しない理由は、形態素のみの方式に比べ、文の自然さや処理速度で劣るためである。

5　評価実験

5.1　人間による識別性

本実験では、全盲／弱視の視覚障碍者らに参加して頂いた。内訳は、スクリーンリーダなどの音声による情報取得者 11 人（U_A）と、拡大鏡などを利用した視覚による情報取得者 8 人（U_V）で、全員が日本語を母国語とする。実験用プログラムは、表 1 の素材文章 B を用いて、1 文あたり 40 文字程度のワードサラダを事前に生成し、それを質問形式にして被験者に提示した。

実験 1.

まず、固定階数の N 階ワードサラダで、人間が感ずる文の自然さが有意となる階数を評価する。6 文からなる $N = 1, 2, 3$ 階ワードサラダを生成し、それぞれの自然さを「5 : 自然」–「1 : 不自然」の 5 段階で評価して頂いた。評価値の（標本平均, 標本分散）は、$N = 1$ で (2.2, 1.4)、$N = 2$ で (3.0, 2.3)、$N = 3$ で (4.4, 1.0) であった。付録 A の検定による $|D|$ 値は、$N = 1, 2$ の比較で $|D| = 0$、$N = 2, 3$ の比較で $|D| = 16$、$N = 1, 3$ の比較で $|D| = 14$ であった。よって、被験者らによる $N = 1, 2$ と $N = 3$ のワードサラダに対する文の自然さの識別は有意である。

実験 2.

次に、[2, 3], [2, 4] 階ワードサラダと 1 階ワードサラダを、人間が識別できるかを評価する。[2, 3], [2, 4] 階ワードサラダを 5 文ずつ生成し、各文について実験 1 と同じ評価をして頂いた。評価値の（標本平均, 標本分散）は、$N = [2, 3]$ で (3.6, 2.1)、$N = [2, 4]$ で (3.6, 2.0) であった。付録 A の方式で、実験 1 の 1 階ワードサラダと [2, 3], [2, 4] 階ワードサラダ各文を検定すると、$N = [2, 3]$ では 4/5、$N = [2, 4]$ では 5/5 の文で有意であっった。したがって、1 階ワードサラダを不自然な文とし、[2, 3], [2, 4] 階ワードサラダを自然な文として扱えば、人間は文の不自然さを有意に識別でき、不自然な文の識別問題による CAPTCHA の実現が期待できる。

被験者群 U_A と U_V の有意性については、結果のみ示す。有意水準 5% での U 検定では、これらの群の間に有意性はないと検定された。

5.2　機械検索結果の評価

Google ウェブ検索を用いて、実験用プログラムが生成したワードサラダの検索結果を評価する。評価値は、「-1 : 文章源が第一候補として検出された」、「0 : 文章源は第二候補以降だが、最初のページに検出された」、「1 : 文章源は特定されない、または第二ページ以降で検出された」とする。等分散を仮定した有意水準 5% の t 検定結果を、表 2 に示す。標本数は、全ての階数で 50 とした。

まず、節 5.1 で使用した素材文章 B から生成したワードサラダについて、表 2(a) を確認する。1 階ワードサラダと [2, 3], [2, 4] 階ワードサラダの検索結果に有意性はないが、文章源が特定される場合が見受けられた。文章源を特定できれ

表 2: ワードサラダの検索結果の評価

(a) 素材文章 B† から生成したワードサラダ

	$N = 1$	$N = [2, 3]$	$N = [2, 4]$	$N = [3, 4]$	$N = 4$
標本平均	-0.32	-0.58	-0.34	-0.48	-0.44
t 値	—	1.65	0.12	0.99	0.76
$N = 1$ との有意性	—	なし	なし	なし	なし

(b) 素材文章 A–E‡ から生成したワードサラダ

	$N = 1$	$N = [2, 3]$	$N = [2, 4]$	$N = [3, 4]$	$N = 4$
標本平均	0.86	0.58	0.84	0.20	0.10
t 値	—	2.13	-0.21	4.29	-5.09
$N = 1$ との有意性	—	なし	なし	あり	あり

† : 表 1 の素材文章 B と同一
‡ : 表 1 の素材文章 A–E を合わせたもの

表 3: 10000 回の試行による新規文の生成確率 [%]

素材文章 † (文字数, 行数)	$N = [1, 3]$	$N = [1, 4]$	$N = [2, 3]$	$N = [2, 4]$	$N = 4$
A (2407,11)	99.84	99.63	56.48	62.93	6.93
B (87260,1717)	100.00	100.00	99.95	99.98	93.23
C (181026,2971)	100.00	100.00	99.99	99.99	95.16
A–E (380783,5248)	100.00	100.00	100.00	100.00	98.2

† : 表 1 と同じもの。A–E は、これら 5 つの文章を合わせたもの

ば、攻撃者はその形態素 N-gram の分布を調査し、回答のヒントにできてしまう。

対策として、複数種類の文書をまとめて素材文章とした結果を、表 2(b) に示す。1 階ワードサラダと [2, 4] 階ワードサラダの検索結果に有意性はなく、かつ文章源の特定確率も 10% 以下となった。文章源が特定されるようなワードサラダについては、事前にフィルタリングし、作問には使用しないようにする。

$N = 1, [2, 4]$ と $N = 4$ の検索結果が異なることから、複数階数を用いたワードサラダ生成法の有用性がわかる。また、$N = [2, 4]$ と $N = [2, 3], [3, 4]$ では、前者の方が $N = 1$ の結果に近いことから、品詞・活用形を用いたマルコフ連鎖の効果も確認できる。

多様な素材文章からワードサラダを生成しても、節 5.1 の結果が変わられなければ、提案手法は識別要件を満たすと考えられる。この点の再評価は、今後の課題となる。

5.3 問題文新規性の評価

問題文としての新規性の評価には、生成されるワードサラダが過去に生成されたものと一致しないことを確認する。表 3 に結果を示す。A–E をまとめて素材文章とした場合は、同じワードサラダの再生成はおこなわれない。

節 5.2 の結果から、文章源が特定される作問に不向きなワードサラダの生成確率は 10 % である。よって、生成したワードサラダの 90 % が問題文として使用できるため、提案手法は、問題新規性要件を満たすと考えられる。

6　まとめ

本稿では、セキュリティ技術の特定知覚への依存の問題を取り上げ、代表例である人間ロボット判別テストでの要件を論じ、文意文脈解釈問題による実現の問題点を指摘した。提案方式は、複数階数による形態素のマルコフ連鎖モデルと、品詞・活用形によるマルコフ連鎖モデルを生成する。本モデルを用いることで、人間には比較的自然な文でありながら、新規性が高く検索による攻撃に耐性のあるワードサラダが生成できる。また、実験を通して、提案手法の有用性を示唆した。

今後の課題としては、人間による識別性の再調査と、実際に CAPTCHA として提示した場合の有効性を確認したい。

謝辞

本稿作成にあたり、産業技術総合研究所セキュアシステム研究部門の中田亨様には、終始適切な助言を頂きました。ここに感謝の意を表します。

参考文献

[1] The official captcha site, http://www.captcha.net/.

[2] Jeffrey P. Bigham and Anna C. Cavender. Evaluating existing audio captchas and an interface optimized for non-visual use. CHI '09, pages 1829–1838. ACM, 2009.

[3] Elie Bursztein, Steven Bethard, Celine Fabry, John C. Mitchell, and Dan Jurafsky. How good are humans at solving captchas? a large scale evaluation. SP '10, pages 399–413. IEEE Computer Society, 2010.

[4] Philip K. Dick. Do androids dream of electric sheep? 1968.

[5] Jonathan Holman, Jonathan Lazar, Jinjuan Heidi Feng, and John D'Arcy. Developing usable captchas for blind users. Assets '07, pages 245–246. ACM, 2007.

[6] Taku Kudo and Yuji Matsumoto. Japanese dependency analysis using cascaded chunking. CoNLL'02, pages 63–69, 2002.

[7] Christopher Liam. System and method for delivering a human interactive proof to the visually impaired by means of semantic association of objects, uspto applicaton 20120232907, 2012.

[8] Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford. Captcha: Using hard ai problems for security. volume 2656 of *LNCS*, pages 294–311. Springer-Verlag, 2003.

[9] H. Wimmer and J. Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1):103, 1983.

[10] 山本 匠, J.D. Tygar, and 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, [コンピュータセキュリティ]. 2009(37):1–8, 2009.

[11] 鴨志田 芳典 and 菊池 浩明. 文章合成の不自然さの評価と応用. ファジィシステムシンポジウム講演論文集. 26:1069–1074, 2010.

A　検定方法

1 つの被験者群に対し、異なる N で生成した 2 つのワードサラダから被験者が感ずる文の自然さの有意性を判定するため、本稿が節 5.1 で用いた検定方法を示す。

- 有意水準は、心理実験で慣用である 5% とする。
- 文 A, B に対し、被験者 i が評価した結果を E_{A_i},E_{B_i} とする。$E_{A_i} > E_{B_i}$ ならば $D_i = 1$、$E_{A_i} < E_{B_i}$ ならば $D_i = -1$、$E_{A_i} = E_{B_i}$ ならば $D_i = 0$ とする。D_i について、19 人の被験者について和を求め、$D = \Sigma_{i=0}^{18} D_i$ を得る。
- 帰無仮説として、D_i が正負ランダム、すなわち二項分布すると仮定する。
- 二項分布に従い D の値を調べると、$|D| \geq 9$ となる確率は 3.2%、$|D| \geq 7$ では 8.4% になる。よって、$|D| \geq 9$ ならば有意水準 5% を満たすので、帰無仮説を棄却する。すなわち、有意差ありと判定する。