

RE-003

「意外や意外」回文だ—文章中に知られざる回文はあるか— Are There Any Unknown Palindromes Embedded in Japanese Texts?

佐藤 理 史†
SATOSHI SATO

1. はじめに

古くからある言葉遊びの一つに、回文がある。『図説 ことばあそび遊辞苑』¹⁾では、回文を次のように説明している。

頭(かしら)から読んでも尻から読んでも同じ音で、どちらも無理なく意味が通じる語句や文章を(回文)という。(p301)

この説明は、次のように、2つの条件に分割することができる。

- (1) 頭(かしら)から読んでも尻から読んでも同じ音であること
- (2) 無理なく意味が通じること

本論文では、前者を回文条件、後者を通意条件と呼ぶ。日本語の回文は、これら2つの条件を満たす日本語表現である。

我々は、このような条件を満たす回文を、自動的に生成することに取り組んでいる²⁾。その過程で、次のような疑問が浮かんできた。それは、「文章中には、知られざる回文が隠れているのではないか」という疑問である。言い換えるならば、我々は意図せず回文を生成しているのではないのだろうか、ということである。

長い回文の作成は、日本語を母国語とする我々にとっても、それほど容易な作業ではない。比較的短い回文、たとえば、「トマト」や「那覇の花」であれば、すぐに思いつく。もし、回文作成の基本戦略—短い回文を組み合わせ延ばしていく³⁾—を知っていれば、これらの短い回文から、より長い回文「那覇のトマトの花(なはのとまとはな)」を作ることができよう。さらに、少し無理をすれば、「スイスや那覇のトマトの花、安いす(すいすやなはのとまとはなやすいす)」を作ることもしできる。

このような長い回文が文章中に隠れているとは考えにくい。が、「トマト」や「那覇の花」のような短い回文は確実に文章中に存在する。それでは、いったい、

- (1) どれくらいの長さまでの回文が実際に文章中に存在するのであるだろうか。

その中には、

- (2) 知られざる面白い回文は存在するだろうか。

本論文では、これらの疑問に答えるために行った調査とその結果について報告する。

2. 方 法

2.1 回文条件

回文条件の「同じ音」をどのように解釈するかには、いくつもの立場があるが、かな表記を音とみなす立場が主流のようである。ここでは、機械的に回文候補を抽出するために、本研究で採用する回文条件を厳密に定義する。

まず、日本語文字列 J のかな表記(ひらがな表記)を次のように定義する。

- (1) J に含まれる漢字を、その読み(ひらがな)で置き換える。
- (2) J に含まれるカタカナを、対応するひらがなで置き換える。
- (3) J に含まれる発音しない記号(句読点、鉤括弧など)を削除する。
- (4) 上記の操作を可能な限り適用することによって得られるひらがな文字列を、 J のひらがな表記とする。

上記の定義は、ひらがな表記の例を示したが、カタカナ表記を採用してもよい。その場合は、(2)の置換は逆方向となる。

このようなに定義されるかな表記を用いて、回文条件を以下のように定義する。

定義 「日本語文字列 J が回文条件を満たす」とは、 J のかな表記 K が次の条件を満たすことをいう。

- (1) K の文字長は2以上である。すなわち、 $|K| \geq 2$ (1) が成り立つ。
- (2) K において、次のことが成り立つ。
 - (a) $|K| = 2m + 1$ (奇数) の場合
 $k_i = k_{2m+2-i}$ ($1 \leq i \leq m$) (2)
 - (b) $|K| = 2m$ (偶数) の場合
 $k_i = k_{2m+1-i}$ ($1 \leq i \leq m$) (3)

ここで、 k_i は文字列 K の i 番目の文字を表す。

なお、本論文では、回文の長さを、かな表記 K の長さとして定義する。回文の長さが奇数となる回文を奇数型、偶数となる回文を偶数型と呼ぶ。奇数型の回文では、ちょうど中央の文字が折り返し文字となる。これに対して偶数型の回文は、中央の2つの文字の間が折り返し点となる。

2.2 厳密条件と緩和条件

式(2)(3)の等号は、文字の完全一致と解釈するのが普通である。このように解釈する場合を、厳密条件と名付ける。

† 名古屋大学大学院工学研究科電子情報システム専攻

人間が回文を作成する場合、清音と濁音、半濁音、促音、拗音を同一視することがある¹⁾。すなわち、

- (1) 濁音(たとえば、「ぶ」)や半濁音(「ぷ」)は、清音(「ふ」)と同じとみなす。
- (2) 促音「っ」は「つ」と同じとみなす。
- (3) 拗音「ゃ、ゅ、ょ」は、それぞれ「や、ゆ、よ」と同じとみなす。

文字の一致をこのように拡大解釈する場合を、緩和条件と呼ぶことにする。

なお、文字の一致の解釈をさらに拡大し、以下のような解釈も採用する場合があるが、本研究では、これらは採用しない。

- (1) 文字「を」を「お」と同じとみなす。
- (2) 助詞「へ」を「え」と同じとみなす。
- (3) 助詞「は」を「わ」と同じとみなす。

2.3 回文抽出法

我々が見つけたいものは、単に回文条件を満たす文字列ではなく、同時に通意条件も満たす日本語の回文である。これを見つけるために、次のような方法を採用する。

- (1) 形態素解析済コーパスを用意する。
- (2) 次の条件を満たす形態素列 M を抽出する。
 - (a) 形態素列 M の形態素数は2以上である。
 - (b) 形態素列 M の読みを連結した文字列 Y は、回文条件を満たす。但し、読みの連結において、読点「、」は削除する。
 - (c) 文字列 Y の長さは、 L_{min} 以上である。
 - (d) 形態素列 M の先頭の形態素は、文節の先頭となりうる形態素(自立語)である。

形態素解析システムは、句読点などの記号の「読み」として、記号自身を出力するのが普通である。このため、読みを連結した文字列 Y は、必ずしもかな文字列になるとは限らない。しかしながら、回文条件のチェックには、特に不都合はないので、式(2)(3)をそのまま用いる。

本抽出法では、読点「、」だけを文字列 Y から削除する。鉤括弧や括弧なども削除することを検討したが、ゴミを拾うことが多かったため、採用しなかった。

上記の方法で、文字列 Y の長さに L_{min} 以上という制限を設けるのは、文字長が短い回文は多数あると考えられるからである。我々の興味は、ある程度以上の長さの回文が存在するか、という点にある。

最後の条件(d)は、通意条件を満たす候補を選択的に抽出するための工夫である。このような条件を課さない場合、付属語で始まる回文候補を多数抽出してしまうことになるが、それらのほとんどは、通意条件を満たさない。

3. 実験

前節で述べた方法を用いて、回文候補(回文条件を満たす形態素列)を抽出する実験を行った。実験には、毎日新聞コーパス15年分を用いた。形態素解析には、mecab+ipadicを利用した。mecab+ipadicは読みをカ

表1 抽出された回文候補数と認定した回文数

Y	厳密条件			緩和条件		
	延べ	異なり	認定	延べ	異なり	認定
17	3	3	0	0	0	0
13	2	2	0	0	0	0
11	5	5	1	1	1	1
9	37	32	7	32	21	14
7	564	247	81	684	264	136
5	50081	1840	-	65204	1582	-
奇数型	-	-	89	-	-	151
12	2	2	0	0	0	0
10	0	0	0	2	2	1
8	18	8	2	32	8	6
6	599	117	21	258	96	37
4	1017	218	-	1096	204	-
偶数型	-	-	23	-	-	44
合計	-	-	112	-	-	195

タカナで出力するため、回文条件判定で使用するかな表記としてカタカナ表記を採用した。 $L_{min} = 4$ として、長さが4以上の回文候補をすべて抽出した。

抽出された回文候補の数を表1に示す。ただし、この表において、緩和条件の欄は、厳密条件には該当しないもののみの数を示している。この表の「異なり」は、日本語形態素列としての異なり数を表す(読みを連結した文字列 Y の異なり数ではない)。

2.3節の条件(d)を満たすからといって、必ずしも通意条件を満たすとは限らない。さらに、形態素解析結果には誤りが存在する。そこで、長さが6以上の回文候補(異なり)をすべてチェックし、回文として認められると著者が判断したもののみを抜き出した。当然のことながら、この判断は人によって揺れるが、本研究の目的は、抽出精度を問うことではないため、複数の人間で判断することはしなかった。

除外した回文候補は、つぎのようなものである。

- (1) 形態素解析の分割誤り、および、読み付与の誤りに起因する誤抽出
- (2) 通意条件を満たさないもの(たとえば、「立った人 事だった」)
- (3) 意味が通じないわけではないが、同じ語の繰り返しなどで、意味が希薄であるもの(たとえば、「ドンドンドンと」、「母はしばしば母」など)

なお、この過程で、新聞記事中に、明らかに回文例として記述されているものも、本調査の趣旨からはずれるため、除外した。たとえば、回文候補には「盛岡の巷、下町の香りも(もりおかのちまたしたまちのかおりも)」が含まれていたが、これは、村上春樹の『またたび浴びたタマ』⁴⁾の紹介記事において、その本に含まれる回文として引用されていたため、除外した。

表1の「認定」欄に、最終的に得られた回文数を示す。得られた307個の回文のすべてを表2に示す。

4. 検討

今回見つかった回文の中で最も長かったのは、次の2

つの回文である。

ないな、反論はないな

(ないなはんろんはないな：長さ 11、厳密条件)

活発か不活発か

(かっぱつかふかっぱつか：長さ 11、緩和条件)

このうち前者は、次の文から抽出されたものである。

で、第二電電がうまくいった時、幹部連中を集めて「みてみい、これで文句はないな、反論はないな」と言いました。(1995年)

上記は、奇数型の回文であるが、偶数型の回文で最も長かったのは、次の回文である。

関係で提言が

(かんけいでていげんが：長さ 10、緩和条件)

これらの例から、長さが 10 を越える回文が文章中に隠れていることが実証された。言い換えるならば、「文章中には、知られざる回文が隠されている可能性がある」ということである。しかしながら、新聞 15 年分のテキストにおいて、長さが 10 を越える回文はたった 3 例であり、このような長い回文が文章中に多数隠されているということはなさそうである。実際、回文候補として抽出された中で長さが 10 を越え、かつ、通意条件を満たすものは、上記の 3 例を除いてすべて意図的に作成された回文であった。このことから、10 を越えたあたりに、無意識的に作られ得る回文の長さの限界があると予想される。

奇数型と偶数型では、明らかに奇数型の方が数が多い。たとえば、長さ 7 の回文は、 $81+136=217$ 個見つかったが、それよりも短い長さ 6 の回文は、 $21+37=58$ 個しか見つからなかった。表 1 の結果に基づけば、奇数型と偶数型の回文数の比は、およそ 7 対 2 である。偶数型では、折り返し点の前後で同じ文字の連続が必要である。このことが、偶数型回文作成の難易度を上げていると考えられる。

文字の一致の条件を緩和すると、より多くの表現が回文と認定されることになる。緩和条件で新たに認定される回文の数 (195 個) は、厳密条件で認定される数 (112 個) の約 1.7 倍である。この事実は、条件を緩めた方が回文が作りやすいという、我々の素朴な直感を裏付ける。

見つかった回文の具体例に目を向けると、なかなか面白い回文が見つかっている。長さ 9、長さ 8 の回文では、次のようなものが秀逸であろう。

- (1) 異端にも忍耐 (いたんにもにんたい)
- (2) 食いつつ見つつ行く (くいっつみっついく)
- (3) たいてい聞いていた (たいていきいていた)
- (4) 断然安全だ (だんぜんあんぜんだ)
- (5) 立て看板が出た (たてかんばんがでた)
- (6) 腹部から下腹部 (ふくぶからかふくぶ)
- (7) 移管委員会 (いかんいいんかい)
- (8) 立ったままだった (たつたままだった)

この実験を通して、著者は、これまで回文と認識できていなかった回文を、多数発見した。意外や意外、「意外

や意外 (いがいやいがい)」が回文であった。この手のパターンは、他に、「異例も異例 (いれいもいれい)」がある。我々は、通常、これらの表現を漢字かな交じりの表記で認識しているため、指摘されなにかぎり、回文と気付かないことが多い。「臨界管理 (りんかいかんり)」や「西が東に (にしがひがしに)」も、見つけたときはかなり驚いた。

これらは厳密条件を満たす回文であるが、緩和条件の場合は、さらに思いもよらぬものが多数あった。「うどん、関東 (うどんかんとく)」、「居酒屋火災 (いざかやかさい)」、「食べ方が下手 (たべかたがへた)」などが回文であることに、これまで気が付いていた人が何人いたであろうか。7 文字の回文は、まだまだ面白いものが文章中に隠れていそうである。

長さ 5、長さ 4 については、抽出された回文候補の数が多かったため、今回は認定対象からはずした。しかしながら、短いからといって一概につまらないとは限らず、その中でも、面白い回文はいくつもあるようである。目についたものを以下に示す。

- (1) 今、うまい (いまうまい)
- (2) 消える駅 (きえるえき)
- (3) 敷いた石 (しいたいし)
- (4) 滝に来た (たきにきた)
- (5) 手でなでて (てでなでて)
- (6) ノリノリの (のりのりの)
- (7) 募金規模 (ぼきんきぼ)
- (8) 自己誇示 (じここじ)

5. 結 論

本研究では、新聞 15 年分のテキストを対象に、文章中に含まれる回文の調査を行なった。得られた結果は、以下のとおりである。

- (1) 文章中に、長さ 10 を越える回文が存在した。
- (2) 意図せずに作られ得る回文の長さの限界は、10 を越えたあたりにありそうである。
- (3) 長さ 9、長さ 8 の回文は、数は少ないが、秀逸なものが見つかった。
- (4) 長さ 7、長さ 6 の回文は、文章中に多数存在した。その中には、知られざる面白い回文が存在した。

謝辞 本研究では、CD-毎日新聞データ集 (1991 版-2005 版) を使用した。

参 考 文 献

- 1) 荻生待也 (編著): 図説 ことばあそび遊辞苑, 遊子館 (2007).
- 2) 鈴木啓輔, 佐藤理史: 文節結合による回文の自動生成, 2010 年度人工知能学会全国大会論文集 (第 24 回), 3D4-3 (2010).
- 3) 土屋耕一: 軽い機敏な仔猫何匹いるか, 角川文庫 (1986).
- 4) 村上春樹: またたび浴びた夕マ, 文藝春秋 (2000).

表2 見つかった回文 (307個)

末尾の数字は回文長、*は緩和条件を表す

ないな、反論はないな/11	クスリのリスク/7	医大で遺体/7*	女兒の所持/7*	善博秘書/7*
活発か不活発か/11*	クスリはリスク/7	伊良部はフライ/7*	女兒や女子/7*	いい父、いい/6
関係で提言が/10*	くすりはリスク/7	動いていこう/7*	じらしにじらし/7*	いい母いい/6
異端にも忍耐/9	この子はこの子/7	動きは気候/7*	高い映画だ/7*	いい母、いい/6
食いつつ見つつ行く/9	この夫婦の子/7	薄暗く、数/7*	高い新潟/7*	金に人気/6
たいてい聞いていた/9	小南奈美子/7	うどん、関東/7*	戦い方だ/7*	小池恵子/6
断然安全だ/9	ゴンタがタンゴ/7	お帰り、笑顔/7*	たたみ込み、ただ/7*	小池桂子/6
人間宣言に/9	再編閉鎖/7	絵画が開花/7*	立ったのだった/7*	子供も、どこ/6
犯人三人は/9	参加し、監査/7	開山祭が/7*	絶ったのだった/7*	子供もどこ/6
犯人3人は/9	傘下の監査/7	改善成果/7*	たった二つだ/7*	この父の子/6
遺伝学鑑定/9*	清水泉氏/7	買った時代が/7*	立つか、分かった/7*	この母の子/6
会談本体が/9*	所持の女子/7	買う方法が/7*	立つためだった/7*	さっき喫茶/6
開発失敗が/9*	処女女子/7	数あるアスカ/7*	食べ方が下手/7*	女子、所持/6
館員全員が/9*	印を記し/7	かつては鉄が/7*	食べ方がへた/7*	対大分/6
歓迎の意見が/9*	人事課人事/7	カナダで田中/7*	試し始めた/7*	第七位だ/6
漢字や写真家/9*	人事部人事/7	カリブ、アフリカ/7*	胆管がんだ/7*	銅の濃度/6
官邸の移転が/9*	世界へ生かせ/7	関係県が/7*	丹沢サンタ/7*	2回以下に/6
参事官が審査/9*	世界へ生かせ/7	寛斎さんが/7*	大体大だ/7*	巻いた大麻/6
立て看板が出た/9*	大差の最多/7	管財傘下/7*	脱退だった/7*	皆は花見/6
球好みの駒田/9*	たかしのしかた/7	神崎さんが/7*	賃貸団地/7*	めどに2度目/6
鉄サクは腐って/9*	互いに鋳型/7	漢詩詩人が/7*	つがい生活/7*	焼いた鯛や/6
日本沈没に/9*	たけのこのけた/7	韓氏自身が/7*	続いて五つ/7*	若い井川/6
配達手伝いは/9*	竹屋が焼けた/7	菅氏自身が/7*	妻が高松/7*	飯田大尉/6*
腹部から下腹部/9*	谷川が似た/7	感謝心が/7*	定価以外で/7*	いかに苦い/6*
移管委員会/8	球の猪俣/7	幹事自身が/7*	適した時期で/7*	意義大きい/6*
父の母の父/8	代打の代打/7	幹事の真価/7*	撤退だって/7*	意義、大きい/6*
意見が関係/8*	地域的位置/7	蓋然性が/7*	手伝いだって/7*	田舎がない/6*
遺産が関西/8*	父の日の父/7	がい然性が/7*	てつだいだって/7*	岡田孝生/6*
立ったままだった/8*	定価を書いて/7	気功の動き/7*	徹夜でやって/7*	岡田孝雄/6*
帝が戦いで/8*	天狗の軍手/7	機構の動き/7*	点差は3で/7*	岡田孝夫/6*
展開いかんで/8*	投与しようとして/7	気功は動き/7*	できずに過ぎて/7*	岡田隆男/6*
母親や叔母は/8*	土井さんサイド/7	機構も動き/7*	出てすぐ捨てて/7*	岡田隆雄/6*
いい年といい/7	土井さん、再度/7	期待を抱き/7*	統合、高度/7*	岡田隆郎/6*
いい夫婦、いい/7	二岡の顔に/7	喫茶の雑記/7*	投与量と/7*	岡田隆夫/6*
いかない仲居/7	西が東に/7	きては子は出来/7*	突然セット/7*	奥田拓男/6*
怒りは理解/7	認定店に/7	禁止の審議/7*	飛び立った人/7*	ガスがかすか/6*
怒りを理解/7	脳の機能の/7	禁止、無人機/7*	飛び出した人/7*	禁止、審議/6*
井川は若い/7	初献血は/7	議席、2議席/7*	飛び出した人/7*	鯨らしく/6*
いかに展開/7	母の日の母/7	黒木の記録/7*	豊科署と/7*	結果が告げ/6*
意外に意外/7	原氏に白羽/7	黒地に白く/7*	どう思おうと/7*	言葉はどこ/6*
意外も意外/7	まさに何様/7	再発売さ/7*	同高校と/7*	言葉は、どこ/6*
意外や意外/7	やすいスイスや/7	再販売さ/7*	どうも妹/7*	五所美子/6*
移籍を規制/7	柳川、金屋/7	魚の長さ/7*	流すわずかな/7*	三次審査/6*
痛いよ、痛い/7	山内厩/7	サワ氏の仕業/7*	2子が利かしの/7*	3次審査/6*
痛いよ痛い/7	理解に怒り/7	3月、監査/7*	二次的で詩に/7*	参事審査/6*
移転、官邸/7	力士が仕切り/7	3次の審査/7*	林記者は/7*	穴戸利二/6*
移転、算定/7	力士は仕切り/7	3次の審査/7*	反撃権は/7*	仕立てて出し/6*
異例も異例/7	臨界管理/7	しっかり活字/7*	半沢さんは/7*	審査参事/6*
会見、経過/7	68は黒/7	知り、がっかりし/7*	半澤さんは/7*	死んだ男児/6*
解析成果/7	若い相川/7	進化も感じ/7*	バッグと靴は/7*	人事信じ/6*
核か非核か/7	輪になれ浪花/7	進化を感じ/7*	バッグや靴は/7*	点差3で/6*
鑑査で参加/7	意見歓迎/7*	真価を感じ/7*	腹部、下腹部/7*	殿下、がんで/6*
監査に参加/7	移行に動い/7*	深化を感じ/7*	ふと見笑み飛ぶ/7*	糖の濃度/6*
菅氏の真価/7	居酒屋火災/7*	侵害監視/7*	ふんした新婦/7*	2ケタだけに/6*
漢詩の真価/7	依然、軍政/7*	紳士と信じ/7*	米団体へ/7*	軒続きの/6*
かんたん短歌/7	以前、県政/7*	信じて人事/7*	ベトナムなどへ/7*	版画館は/6*
感度の鈍化/7	以前、先生/7*	診断、男児/7*	三つに包み/7*	バッグ、靴は/6*
幹部は、文科/7	以前、男性/7*	新婦にふんし/7*	見て作ってみ/7*	目当てでアメ/6*
軌跡と奇跡/7	以前、民生/7*	針路を論じ/7*	門外漢も/7*	理屈作り/6*
禁煙延期/7	委託拡大/7*	次回に生かし/7*	慶喜の書/7*	
禁止の新規/7	いただいた鯛/7*	時間変化し/7*	義珍秘書/7*	
菓のリスク/7	抱いていた/7*	女児と女子/7*	義仁秘書/7*	