

近代語テキストからの可能動詞の抽出

「太陽コーパス」を例に

小木曾 智信

キーワード：近代語，可能動詞，電子化テキスト，検索方法

はじめに

近年、日本語研究においてもコンピュータの利用が進み、大規模なデータを用いた用例ベースの研究も増えつつある。とはいえ、現在利用可能なデータの多くはプレーンテキストを中心とする、文字列検索のみが可能なデータであって、語ごとに区切られ品詞情報などが付与されたものは多くない。形態素解析システムの利用なども考えられるが、少なくとも現代語以外の資料に対してはこれも無力であり、結局は文字列検索に頼ることになる。

したがって、意図した検索対象の語が、他の文字列から区別されるに十分な長さを持つ文字列として表記される場合や、他の語では用いられないことのない低頻度の文字を含む場合を除き、いわゆる「ゴミ取り」作業、すなわち検索結果から意図したものと異なる例を排除してゆく作業が重要となる。場合によっては、この作業がデータ処理の最も多くの部分を占めるとしてもよいかもしれない。しかし、最終的に利用されるのは処理済みの用例データのみであり、用例抽出作業の方法そのものについて触れられることはないのが普通である。こうした作業は、表に出すべきではない影の作業と位置づけられているように思われる。

しかし見方を変えれば、テキストの表層情報から必要とする情報であるかどうかを判別するという作業は日本語研究者であるからこそ可能な高度な作業である。そのノウハウが埋もれ、忘れ去られてゆくのは残念なことである。

また一方で、今後データ量がよりいっそう増えてゆくに従い、こうした手法そのものをより前面に押し出してくることが必要になるのではないかと思われる。処理しきれないほどの検索結果からどのようにして用例を抜き出したのかが明示されなければ追試も不可能である。その抜き出し処理の方法・語の検出アルゴリズムそのものも、論考の元となった資料の確実性、ひいては研究の信頼性を裏付けるための基礎的な情報として重要なものとなってくると思われる。

以上のような観点から、本稿では近代語のテキストから可能動詞を抽出する方法について考える。可能動詞の使用が急拡大し助動詞「れる」による可能表現を駆逐してゆくようになるのは近代以降のことであるため近代語の調査として興味深いのみならず、その検索が一筋縄では行かないことから、検索方法そのものを考える必要があるからである。

対象とするテキスト

国立国語研究所で開発中の『太陽コーパス』を対象とした。これは他に類を見ない大規模な近代日本語のテキストデータであり、多彩なジャンル・文体を網羅していることから当時の言葉の実態を探るのに格好のものである。ただし、以下で述べる検索の方法は、これ以外の近代語のテキスト、さらには現代語・近世語のテキストにも適用できるものを目指しており、これに特化したものとはしない。

検索の方針

可能動詞そのものを特徴づける形態的な特徴は、全てに共通する文字列にすると「工段の仮名」ということにしかならない。これだけでは到底検索語としての用をなさない。そこで、ありうる可能動詞の語形リスト（辞書）を作り、これを検索語として検索を行い、その結果に対して絞り込みをかけるという方法を取ることにした。

可能動詞はきわめて生産的であり、一部の無意志自動詞などを除き、ほとんどの五段動詞から生成される。そこで、ひとまず可能動詞は五段活用動詞から自由に作られると仮定して、国語辞典から五段活用動詞のリストを作り、これを機械的に可能動詞の形に変形することにした。

これらの一連の作業や検索システムの実装は、主に自作のperlスクリプトによって行った。

表記のバリエーション

ところで、日本語のテキスト、とりわけ近代語テキストにおける語の検索で障害となるのは表記の多様性である。『太陽』本文の様態も田中・小木曾（2000）にみるように実に多様であって、可能動詞の検索に際しても、漢字・仮名遣いなど表記法のさまざまな部分で問題がでてくる。これについては次のような方針で対処した。

漢字の異表記

漢字の異表記については、国語辞典（『岩波国語辞典 第五版』）所載の異表記全てを候補とした。例えば「ひける<ひく」の場合には「引・弾・曳・牽・挽・碾・轆・退」である。

こうした語のレベルでの異表記に加え、旧字・異体字のような文字レベルでの異表記も存在する。これについては一般的な新字・旧字の対応表によって両方の漢字を候補とした。例えば上にあげた「弾」に対する「彈」などである。

仮名遣いの変異

原文では歴史的仮名遣いに一致しない例が多いが、『太陽コーパス』では修正注が付けたうえで、仮名遣いを全て歴史的仮名遣いに統一している。このほかのテキストについても文学作品など多くのデータは仮名遣いが整備されたものであることが期待できる。そうしたこともあり、今回は現代仮名遣いと歴史的仮名遣いの両方を候補としたが、そのいずれにも一致しない仮名遣いは候補としなかった。例えば、「食える」「食べる」は検索するが、仮に「食ゑる」という形があってもこれは検索していない。ただし「っ」などの小書き片仮名については「つ」「っ」の両方を候補とした。

濁点

原文に濁点が付けられていない場合も多い。『太陽コーパス』では原則として濁点を付し修正注が付けられているが、今回は見出し語に濁点が含まれる場合には濁点なしのものも候補とした。

送り仮名の処理

今日の用法から見て、送り仮名が極端に省略されているような場合がある。例えば「書た」とあってもこれが「書けた」なのか「書いた」なのか判断することはできない。原文でルビが付いている場合には、別途ルビを開いたテキスト（親字の代わりに振り仮名文字列を埋め込んだテキスト）を用意してこれを検索することとし、通常のテキストではこうしたものの検索は見送った。ただし国語辞典中でも送りがなの揺れが認められているような一般的な例については両方の形が検索できるように配慮した。

踊り字の処理

踊り字が用いられる場合も多い。これについては検索システムの側ではなく、テキストの側をあらかじめ処理することで対処した。くの子点が影響することは考えにくいため、「ゝ」「ゞ」などの一字を繰り返すものについてのみ、直前の仮名（またはその濁音仮名）に置き換える操作を行った。

検索語リストの作成

以上をふまえ、次のような手順で検索語のリストを作成した。

まず、『岩波国語辞典 第五版』所載の見出し語のうち、〔五他〕〔五自〕〔五自他〕の記述がある語のリストを作る（2701語）。この見出しの漢字表記それぞれについて、旧字・異体字などの文字レベルでの異表記がある場合にはこれを新たに加えた。

ここから複合動詞などの、後方からみて文字列が重複することになるものを取り除き、必要な五段動詞のリストを作成した。この重複削除は、仮名と漢字で別々に行う必要がある。例えば「さかのぼる【遡る】」は、仮名では「のぼる」と重複するため削除するが、漢字では削除しない。

この操作によって、漢字表記総計で1723、仮名表記で755のリストができる。この五段動詞リストのそれぞれの語について、活用の行にあわせて語形を工段に変更することで可能動詞の語幹リストを作成する。

一段活用動詞と同形となる場合

このようにして五段活用動詞から機械的に作り出した可能動詞形が、同一語幹の一段活用動詞と同形になる場合がある。その中には、例えば「遅らせる<遅らす」「震える<震う」のように、可能動詞として用いられる場合がないものも多く含まれている。

これらを除外するために、これまでに絞り込んだ可能動詞候補のリストを、『岩波国語』から抽出した一段活用動詞のリストと対照し、漢字見出しが一致するものをはじき出した。仮名についてこれを行うと「あえる」（会える・和える）、「いえる」（言える・癒える）など別語が除外されてしまうため、漢字見出しについてのみこの処理を行った。次ページの表に示すとおり、このように同形となる場合が漢字見出しで157ある。

可能動詞形と同形となる一段動詞（漢字表記）

うなずける	額れる	×	あからめる	赤らめる	×	そろえる	揃える
うれる	売れる	×	あける	明けれる	×	たいらげる	平らげる
おける	置ける	×	あわせる	合(わ)せる	×	たがえる	違える
おれる	折れる	×	いそがせる	急がせる	×	たわめる	撓める
かまえる	構える	×	いただける	頂ける	×	ちぢめる	縮める
きれる	切れる	×	いただける	戴ける	×	つがえる	番える
くえる	食える	×	いためる	痛める	×	つからせる	疲らせる
くだける	砕ける	×	いためる	傷める	×	つける	漬ける
くつろげる	寛げる	×	うつぶせる	俯せる	×	つける	浸ける
さける	裂ける	×	うつむける	俯ける	×	つける	附ける
さばける	捌ける	×	おくらせる	後らせる	×	つける	付ける
しれる	知れる	×	おくらせる	遅らせる	×	つたえる	伝える
すすめる	進める	×	かがめる	屈める	×	つづける	続ける
すれる	摺れる	×	かける	欠ける	×	つめる	詰める
すれる	擦れる	×	かしげる	傾げる	×	てるる	照れる
すれる	磨れる	×	かすれる	掠れる	×	とける	溶ける
すれる	揺れる	×	かたぶける	傾ける	×	どける	退ける
すれる	刷れる	×	かなえる	叶える	×	とどける	届ける
すれる	摩れる	×	かなえる	適える	×	ととのえる	整える
そげる	削げる	×	からめる	絡める	×	ととのえる	調える
そげる	殺げる	×	きかせる	聞かせる	×	なぐさめる	慰める
たてる	立てる	×	きかせる	利かせる	×	におわせる	匂わせる
たのしめる	楽しめる	×	きずつける	傷つける	×	におわせる	臭わせる
つける	着ける	×	くくめる	銜める	×	にぎわせる	賑わせる
つける	就ける	×	くされる	腐れる	×	ねかせる	寝かせる
つなげる	繋げる	×	くじける	挫ける	×	ねじれる	捻(じ)れる
つれる	釣れる	×	くだされる	下される	×	ねじれる	捻(じ)れる
つれる	攀れる	×	くびれる	縊れる	×	のぼせる	上せる
つれる	吊れる	×	くぼめる	窪める	×	ひしげる	拉げる
とける	解ける	×	くわせる	食らわせる	×	ひそめる	潜める
とれる	捕れる	×	くるしめる	苦しめる	×	ひびかせる	響かせる
とれる	取れる	×	狂わせる	狂わせる	×	ふくめる	含める
なける	泣ける	×	くろめる	黒める	×	ふせる	伏せる
ぬける	抜ける	×	くわせる	食わせる	×	ふせる	臥せる
ぬげる	脱げる	×	こめる	込める	×	ふるえる	震える
ねれる	練れる	×	さわがせる	騒がせる	×	ふるわせる	震わせる
はげる	剥げる	×	しずめる	沈める	×	まがえる	紛える
はなせる	話せる	×	したがえる	従える	×	まかせせる	任せせる
ひける	引ける	×	しのばせる	忍ばせる	×	まぎらせる	紛らせる
ふるる	振れる	×	しらせる	知らせる	×	まくれる	捲れる
ほれる	掘れる	×	しわめる	皺める	×	またげる	跨げる
むける	剥ける	×	すがめる	眇める	×	もめる	揉める
むける	向ける	×	すくめる	竦める	×	もれる	洩れる
やける	焼ける	×	すける	透ける	×	もれる	漏れる
やぶける	破ける	×	すたれる	廢れる	×	やすめる	休める
よれる	繕れる	×	すぼめる	窄める	×	やめる	止める
われる	破れる	×	すませる	済ませる	×	やめる	病める
われる	割れる	×	せかせる	急かせる	×	やめる	已める
よめる	読める	×	そえる	添える	×	やわらげる	和らげる
あける	開ける	×	そえる	副える	×	ゆるめる	緩める
いれる	入れる	×	そだてる	育てる	×	ゆるる	揺れる
ならべる	並べる	×	そむける	背ける			
× あおのける	仰げる	×	そめる	染める			

(異体字などは含めず、現代仮名遣いでの形のみを挙げた)

この中には可能動詞として使用されるものも含まれている。表中、**印**を付けたものは可能動詞としての用法も認められると考え、リストに残した。

同じ動詞でも表記によって使い分けがある場合には、それぞれの使用状況に即してリストに残すかどうかを決めた。例えば「つける」は「着ける」「就ける」は残し、「漬ける」「浸ける」などは除いた。「とける」では「解ける」を残し「溶ける」「熔ける」などは除いた。

可能動詞候補から除外されるものを簡単に整理すると、多くは次のようなものである。

- ・ 元から五段活用・一段活用の間で揺れがあるもの、またはほぼ同義で両方の活用があるもの。

例：「震える<震う」「漏れる<漏る」「洩れる<洩る」

- ・ 上のうち、特に「せる」と「す」が交代するもの。

例：「聞かせる・聞かす」「狂わせる・狂わす」「忍ばせる<忍ばす」

- ・ 身体動作を表す自動詞の可能動詞形が、身体部分を対象とする再帰的な動詞と一致するもの

例：「俯ける・俯く」「屈める・屈む」「背ける<背く」

- ・ 無意志的な自動詞、非対格自動詞から作られた可能動詞形が、一段活用の他動詞と一致するもの

例：「赤らめる<赤らむ」「叶える<叶う」「育てる<育つ」「揃える<揃う」「縮める<縮む」「続ける<続く」「届ける<届く」「整える<整う」「止める<止む」

- ・ 元となる五段動詞が意志的な動詞であっても可能動詞形が考えにくいもの

例：「伝える<伝う」「捲れる<捲る」「揉める<揉む」「擦れる<捻る」

「苦しめる」「沈める」なども、意志動詞として「苦しむ」「沈む」が用いられる場合には可能動詞としての用法も考えられるわけだが、そのような実例は考えにくいうえに自動詞との誤ヒットが予想されることから除外した。

さらに、ここでは「陥(おちい)れる<陥(おちい)る」に対する「陥(おとしい)れる」、「貶(さげす)める<貶(さげす)む」に対する「貶(おとし)める」のように、漢字表記上一致するものも可能動詞として用いられない場合には削除した。一方、「入(い)れる」「入(はい)れる」のように表記上一致することがあるために残したものがある。

ここでリストに残すことにしたものの多くは、同音異義語と考えないのであれば、可能動詞としての用法を持ちながら、可能動詞の範囲を超えた用法を持つ動詞ということになる。これらについてはその歴史的な経緯を含め、意味の範囲などをより深く検討する必要があるだろう。ここで排除した動詞についても可能動詞との関わりを含めて考えるべきことは多いが、いずれも今後の課題としたい。

なお、ここであえてリストに残した語は、この語形をしていても可能動詞である可能性が低くなると考えられる。そこで後述する絞り込み時にこの不確かさや構文的な条件を反映することにする。

正規表現

以上の手順で作成されたリストをそのまま検索語として総当たりで試しても良いのだが、実際の検索に膨大な時間がかかるため、次の表のように活用の行ごとに正規表現にまとめて検索語リストとした。ここには漢字見出しで終止形が3文字になるもののみ挙げる。

実際に検索してみると、終止形が4文字以上になる場合には可能動詞に正しくヒットする例はきわ

ある。「書く」を例に取れば、命令形、仮定（已然）形、連用形の「書け」が可能動詞語幹に一致してしまう。しかし、このうち命令形、已然形は直後にあらわれる文字列によってかなり絞り込むことができる。命令形の場合は「！・」・『』などの記号類が来る場合があるが、可能動詞語幹にこれが続く場合はない。また、仮定（已然）形の場合は、係り結びを除き、助詞「ば」「ど」「ども」が来るが、可能動詞語幹にこれが続くことはない。また、文語では助動詞の「り」が続く場合も多いが、これも可能動詞語幹には付かない。このようにして「し・ば・ど・と・り・ざ・ら・を・れ・！・。・、・」・『』などは可能動詞語幹の直後に来ることはないため、除外することができる。

1901年から1925年までの『太陽コーパス』全体を単純に語形のみで検索した場合、前に正規表現の形で挙げた終止形が3文字になる漢字入りの語形だけで約4万件ヒットすることになるが、下接文字列による絞り込みによってこのうち約半数を除去することができる。

上接文字列による絞り込み

可能動詞の直上に来る文字列には、動詞連用形・副詞・名詞句などさまざまなものがありえ、ごく一部の記号類を除けば直接的な制限はかけられない。しかし、実際に可能動詞があらわれる場合には何らかの規則性を持っていると考えられる。

そこで、あらかじめ行った検索結果を手作業で絞り込んで可能動詞の用例リストを作成し、この用例リストにおける可能動詞の直前文字を頻度順で並べたもの（上位20字）が右の表である。

これによると上接字にはかなりの偏りが見られ、「は」「も」「に」「が」「て」などの助詞と思われる文字が来る場合が圧倒的に多いことがわかる。今回は利用できなかったが、この情報を活かすことで絞り込みが可能になるはずである。

「見」「受」が上位にあるのは「見出せる」「受取れる」などの語形を後方一致で採集したためである。また「～」は踊り字（くの字点）を表す。

語別にこの頻度を見ることによって、また、2文字・3文字などこれより長い範囲で頻度の高い文字列を調査することによって、より有効な方法になると考えられる。

可能動詞の直前の文字・頻度表

文字	数	頻度(%)
は	365	18.60
も	298	15.19
に	201	10.25
が	167	8.51
て	139	7.09
と	118	6.01
で	83	4.23
の	62	3.16
を	57	2.91
く	55	2.80
り	49	2.50
へ	30	1.53
か	28	1.43
ら	26	1.33
ひ	24	1.22
し	22	1.12
見	21	1.07
ば	18	0.92
受	11	0.56
～	10	0.51

（上位20字）

文体による絞り込み

比較的新しい語形である可能動詞は、口語で書かれたテキストに多くあらわれ、文語で書かれたテキストには出現しにくいと考えられる。また、最も紛らわしい語形は五段（四段）動詞連用形に助動

詞「り」が付いたものの連体形・連用形であるが、この形は文語文では頻出するのに対し、口語文ではほとんどあらわれない。こうしたことから、文語・口語いずれによって書かれた記事であるのかによって検索結果を分類することで検索結果の絞り込みに生かすことができそうである。

幸いなことに、『太陽コーパス』は記事ごとに[文語][口語][韻文]といった情報がタグ付けされており、これを利用することでこの操作は容易に実現できた。こうした情報がない場合には検索対象指定時に文体を指示するか、文末辞などによる文体の自動判別が必要となるだろう。

この文体による判別は、後述する調査結果からもわかるとおり、かなり有効に機能する。

文中要素による絞り込み

本格的な構文解析を行うことは望むべくもないが、文の中にキーとなる文字列が含まれるか否かによって、ヒットした語の可能動詞としての確からしさを認定することが可能であると思われる。

例えば、他動詞と同形になる「進める」「着ける」の場合、可能動詞と見なされるのは自動詞「進む」「着く」から作られたものであるから文中にヲ格の語はあらわれない。それに対し、同形の一段動詞は他動詞であるためヲ格の語があらわれる場合が多い。したがってヒットした位置より上の文中に文字「を」があらわれない場合は可能動詞である可能性が高いと判断できる。

実際に「進める」の場合を試すと、458ある候補のうち、文中に「を」を含まないものは33例に絞られる。可能動詞として認めた「進める」は次の3例にすぎないが、このいずれもが33例中に含まれていた。用例はいずれもリーヒイ/佐野慶介(訳)の長編科学小説「生ける死」からであった。

- ・ その日はやつと六哩しか進めませんでした。、1925年2号,生ける死(第二回)
- ・ 或る時は一呎も進めなかつた。、1925年7号,生ける死(第六回)
- ・ その日は道が頗る悪かつたので五哩半しか進めなかつた。、1925年10号,生ける死(第八回)

逆に、自動詞と同形となる「破れる」の場合、自動詞「破れる」ではヲ格の語はあらわれず、他動詞「破る」から作られた可能動詞では文中にヲ格の語があらわれることが多くなるようにも思われるが、可能文であるためヲ格がガ格に交代している場合が多く「を」の有無での判別は有効ではない。しかし、これを他動詞から作られる一般の可能動詞に逆に当てはめて、可能動詞であればヲ格がガ格に交代している場合が多いから「を」を含まないものの方が可能動詞である可能性が高い、といった絞り込みも考えられる。

このほかにも利用できる文中要素はさまざまなものがあると思われるが、今後用例を集めた上で検討してゆくことにしたい。

検索結果に対する重み付け

以上述べてきた絞り込みの方法のうち、確実に可能動詞以外のものを排除できると思われるものは下接文字列による絞り込みの一部だけである。この確実なものについては検索プログラムそのものに埋め込む形で実装した。

そのほかの方法は確からしさを増すだけであって、誤って可能動詞を排除してしまう可能性が残る。こうした手法の実装は検索結果に対してフィルタをかけ、確からしさのレベルに応じて重み付けを行い、これを検索結果に反映させるのが望ましい。

この重み付けの方法については試行錯誤を繰り返している段階にある。対象とする資料に合わせてここで調整を行う必要もあるだろう。事例を多く集めた上でなければ十分な効果が期待できないこともあって、今回の検索では部分的に利用したのみである。

『太陽コーパス』における可能動詞の用例数の推移

これまでに述べた検索方法を実際に用いて『太陽コーパス』における可能動詞の用例数の推移を調査した。本調査はいわば上述の方法の実証試験であり、対象とするデータも開発途上の版である。したがって結果についても暫定的なものであることをお断りしておく。

可能動詞の年ごとの用例数を、年ごとのテキスト量・記事の文体別を考慮してまとめると次の表1のようになる。計算に使用した太陽コーパスのテキスト量は表2に示したとおりである。

用例の認定に際し、一段動詞と同形のは「することができる」への置き換えの適否などによって適宜判断した。例えば、「(異性に)持てる」、「(取っ手が)取れる」などは除外し、「(荷物)持てる」、「(連絡が)取れる」などは含めた。また「耐え切れる」「飛び出せる」のような複合動詞の後項になる場合もそれぞれ「切れる」「出せる」のうちに含んでいる。

なお、今回の調査では「知れる」はすべて除外した。幅広い用法を持つうえに「かもしれない」などの形で頻用され、判別が難しいことからである。「切れ味」「売れ行き」のような名詞の一部になっているものなども除いている。

表1 可能動詞の用例数・出現率

年	可能動詞用例数	出現頻度 (語/MB)	口語記事中の 用例数	口語記事中の 出現頻度(語/MB)
1895	90	13.48	39	103.26
1901	166	26.01	134	78.69
1909	412	65.67	386	95.69
1917	531	89.58	488	98.38
1925	762	112.32	759	113.55

表2 太陽コーパスの文体別テキスト量

年	全記事(KB)	文語・ 韻文(KB)	口語・ その他(KB)	口語%
1895	6837	6451	387	5.7
1901	6536	4793	1744	26.7
1909	6424	2294	4131	64.3
1917	6070	991	5080	83.7
1925	6947	102	6845	98.5

(タグなどを除去した正味のテキスト量)

表3 各年の可能動詞異なり語数

年	異なり語数
1895	29
1901	49
1909	71
1917	71
1925	118

可能動詞の用例数・出現頻度(テキスト1メガバイトあたりの出現数)は年を追って増えてゆく。しかし、テキストの口語割合も同じようにして増えているため、単純に増加しているとはいえない。

口語記事に限って見ると、増加の傾向も見て取れなくはないが、それほど急激な増加ではない。

しかし各年の可能動詞の異なり語数を見ると表3のようになる。絶対的な用例数が多いので当然の結果とも言えるが、着実に使用される範囲の幅が広がっている様が見て取れる。実際にあらわれる可能動詞は次の表4に示した通りである。

表4 『太陽』各年の記事にあらわれる可能動詞

1895	愛せる, 引ける, 云へる, 下せる, 割れる, 喰へる, 遣れる, 言へる, 行ける, 持てる, 執れる, 取れる, 出せる, 書ける, 食へる, 吹ける, 切れる, 漕げる, 置ける, 通れる, 登れる, 望める, 貰へる, 乗れる, 弾ける, 飲める, 盡せる, 讀める, 賣れる
1901	逢へる, 謂へる, 云へる, 往ける, 歌へる, 解ける, 居れる, 遣れる, 言へる, 限れる, 悟れる, 行ける, 行へる, 採れる, 作れる, 殺せる, 伺へる, 思へる, 持てる, 取れる, 書ける, 勝てる, 上れる, 食へる, 積める, 切れる, 説ける, 潜れる, 足れる, 置ける, 通れる, 潰せる, 渡せる, 渡れる, 登れる, 買へる, 飛べる, 聞ける, 保てる, 歩ける, 募れる, 眠れる, 貰へる, 遊べる, 抜ける, 拂へる, 讀める, 賣れる, 飛ばせる
1909	愛せる, 逢へる, 謂へる, 云へる, 汲める, 居れる, 喰へる, 遇へる, 撃てる, 遣れる, 言へる, 呼べる, 行ける, 行へる, 採れる, 作れる, 捌ける, 使へる, 思へる, 死ぬる, 飼へる, 持てる, 取れる, 住める, 出せる, 書ける, 勝てる, 笑へる, 食へる, 振れる, 成れる, 切れる, 接げる, 措ける, 足せる, 打てる, 貸せる, 置ける, 通せる, 通れる, 釣れる, 渡れる, 登れる, 働ける, 動ける, 入れる, 買へる, 飛べる, 描ける, 負へる, 保てる, 捕れる, 歩ける, 歩める, 眠れる, 貰へる, 利ける, 練れる, 話せる, 會へる, 描ける, 飲める, 歸せる, 歸れる, 焼ける, 畫ける, 疊める, 盡せる, 讀める, 賣れる, 着こなせる
1917	逢へる, 扱へる, 謂へる, 云へる, 解ける, 掛れる, 吸へる, 泣ける, 居れる, 許せる, 喰へる, 掘れる, 計れる, 結べる, 言へる, 行ける, 行へる, 捌ける, 使へる, 思へる, 死ぬる, 持てる, 借れる, 取れる, 出せる, 書ける, 勝てる, 省ける, 上れる, 食へる, 申せる, 成れる, 切れる, 穿ける, 措ける, 打てる, 貸せる, 置ける, 通れる, 掴める, 釣れる, 渡れる, 働ける, 忍べる, 買へる, 飛べる, 描ける, 負へる, 聞ける, 捕れる, 歩ける, 縫へる, 望める, 眠れる, 貰へる, 遊べる, 頼める, 離せる, 話せる, 乗れる, 會へる, 巻ける, 弾ける, 拂へる, 飲める, 争へる, 盡せる, 讀める, 賣れる, 隠せる, 飛ばせる
1925	愛せる, 逢へる, 握れる, 謂へる, 窺へる, 唄へる, 云へる, 泳げる, 往ける, 果せる, 歌へる, 過せる, 解ける, 願へる, 起せる, 疑へる, 救へる, 泣ける, 去れる, 居れる, 拒める, 許せる, 喰へる, 掘れる, 撃てる, 結べる, 現せる, 言へる, 雇へる, 行ける, 合へる, 作れる, 搾れる, 殺せる, 捌ける, 伺へる, 使へる, 思へる, 施せる, 死ぬる, 持てる, 取れる, 拾へる, 出せる, 書ける, 勝てる, 笑へる, 飾れる, 織れる, 食へる, 申せる, 進める, 推せる, 成れる, 盛れる, 切れる, 説ける, 搔ける, 走れる, 憎める, 造れる, 測れる, 打てる, 置ける, 直せる, 通れる, 掴める, 綴れる, 釣れる, 塗れる, 渡せる, 渡れる, 怒れる, 働ける, 動ける, 忍べる, 覗ける, 買へる, 否める, 被れる, 飛べる, 描ける, 負へる, 貰ける, 聞ける, 歩ける, 歩ける, 暮せる, 放せる, 望める, 防げる, 眠れる, 貰へる, 遊べる, 利ける, 凌げる, 練れる, 話せる, 乗れる, 會へる, 學べる, 弾ける, 戦へる, 抜ける, 拜める, 拂へる, 挟める, 盗める, 飲める, 歸せる, 歸れる, 争へる, 畫ける, 盡せる, 繼げる, 聽ける, 讀める, 賣れる

著者・ジャンルごとの違いや、助動詞「れる」による可能表現など他の形式との対照、文脈の肯定否定など、調査すべきことは多いが、これらのより詳しい検討は別の機会に譲ることとする。

おわりに

今回は使用頻度をほとんど問題にせず、基礎データを国語辞典により、これに記載のあるほとんどの語を候補とした。誤ヒットが多いものや使用頻度が低いものであっても、可能動詞の可能性が残るものは検索語に残したままである。検索語リストについては十分な検討を行っておらず、まだまだ余

分なものが含まれていると考えられる。一方で、実際のテキスト中では、辞書に記載されないような、一般的でない表記がなされる場合も多く、そうしたもののうちに可能動詞を取りこぼしてしまっている可能性は残る。こうしたものは実際の用例を元に辞書に追加してゆかなければならない。

今後、実例をもとに辞書をブラッシュアップし、同時に絞り込みのフィルタについても用例・頻度に基づいた実装を進めて行く予定である。

「太陽コーパス」の形式や規模、雑誌『太陽』本文の様態などについては田中・小木曾（2000）、田中（2001）を参照されたい。同コーパスは現在開発中であるが、本稿執筆者が非常勤研究員として勤務していることから、今回は内部公開されている 版Ver.0.6（2001.12.27版）を利用させていただいた。

「太陽コーパス」の一部は試用版として外部公開されている。これについては田中（2001）付記を参照されたい。

参考文献

- 神田寿美子（1961）「現代東京語の可能表現について」東京女子大学日本文学16（再録：『論集日本語研究15現代語』有精堂1984）
- 鶴岡昭夫（1967）「江戸語・東京語における可能表現の変遷について」言語と文芸54
- 木村睦子（1992）「国定読本における可能表現」『辻村敏樹教授古稀記念 日本語史の諸問題』明治書院
- 渋谷勝己（1993）「日本語可能表現の諸相と発展」大阪大学文学部紀要33第1冊
- 田中牧郎・小木曾智信（2000）「総合雑誌『太陽』の本文の様態と電子化テキスト」『日本語科学』8号
- 田中牧郎（2001）「XMLを利用したコーパスの構築 『太陽コーパス』を中心に」『日本語学』Vol.21.14