

日本語受動文の能動化可否判定アルゴリズムの検討†

林 良 彦** 千葉 裕 子**†*

書換え支援機能を持つ日本文推敲支援システムの研究開発を進めている。その中の推敲支援項目の1つとして、受動文の能動化をとりあげている。文章中に存在する受動文を能動化するには一定の条件があり、すべての受動文を無条件に能動化することはできない。したがって、受動文の能動化を行う際には、まずその受動文が能動化可能であるかを判定する必要がある。本論文で提案する受動文能動化可否判定アルゴリズムは、マニュアル文章を対象とし、能動化できない受動文を文脈上の制約および文内の統語的制約から判定する。文脈上の制約から能動化できない受動文を判定するには、省略されている主語および動作主名詞句の同定が必要である。しかし、マニュアル文章においては、「ユーザ」、「筆者」、「説明対象物」が既知の情報として文章の最初から省略されるため、従来の省略要素同定手法によってこれを同定することができない。そこで、本アルゴリズムでは、マニュアル文章の特性を用いてこれを同定する手法を提案する。評価実験によれば、本アルゴリズムは、能動化が可能な受動文の76.0%、能動化が不可能な受動文の87.2%を正しく判定することができる。

1. はじめに

我々は、書換え支援機能を持つ日本文推敲支援システムの研究開発を進めている¹⁾。その中の1つの推敲支援項目として、受動文の能動文への書換えをとりあげている。これは、受動文が能動文に比べて主張の強さや明確さに欠けることから、マニュアルの文章には好ましくないと言われている²⁾ためである。

しかし、すべての受動文を無条件に能動文に書き換えることができるわけではなく、能動化できる受動文には一定の条件がある。したがって、推敲支援システムにおいては、受動文に対してそれが能動化できるものか否かをまず判定する必要がある。

本論文では、マニュアルの文章を対象として、能動化できない受動文の条件を、文脈上の制約と文内の統語的制約に分けて考察し、これらの制約を用いて判定を行うアルゴリズムを提案する。特に、文脈上の制約による判定を行うには、省略されている主語、動作主名詞句の同定が必要であること、さらに、従来の省略要素同定手法がマニュアル文章においては有効でないことを指摘し、マニュアル文章の特性を利用した省略要素同定手法を提案する。

2. 文脈上の制約から能動化できない受動文

2.1 文脈の制約から能動化できない受動文

文脈の影響で能動化できない受動文がある。例え

ば、以下の例の下線部の受動文は、「機器がデータを自動的に保存する」という意味であるが、それを能動化した文は、「ユーザが前回保存したファイル名をもう一度付けて保存する」という異なった意味を表す。

例：入力したデータはできるだけ頻繁に保存します。メニューから「保存」を選びます。前回保存したファイル名でデータが保存されます。

これは、「文章単位のなかで連続する省略主語は同じものを指すと解釈される」という原則^{3),4)}のために、下線部を能動化した文の省略主語が前文の省略主語と同じく「ユーザ」であると解釈されるからである。

2.2 マニュアルにおける省略要素同定の特徴

以上から、文章単位のなかで連続する省略主語が一定であるという仮定のもとで、ある受動文の能動化可否を判定するためには、前文の主語を同定することおよびその受動文を能動化したときに主語となる受動文の動作主名詞句を同定することが必要である。

実データの分析によれば、マニュアル中の文章は、文章単位ごとにユーザが行うべき操作を説明する文章(操作説明文章)、対象物が何をするか説明する文章(機能説明文章)、筆者がそのマニュアルをどう書いたか説明する文章(マニュアル説明文章)の3つのタイプに分けることができる。また、それぞれの文章において省略主語が指しているものは、原則的に「ユーザ」、「説明対象物」、「筆者」である。ところで、これらの事物はマニュアルという文脈において既知の情報⁵⁾であることから、文章単位の最初から省略されることが多い。したがって、前文から補完する手法や「ハ」で示される主題名詞句で補完するという従来から一般的な省略主語補完の手法を用いることができ

† An Algorithm to Detect Activizable Japanese Passives by YOSHIHIKO HAYASHI and YUKO CHIBA (NTT Communications and Information Processing Laboratories).

** NTT 情報通信処理研究所

* 元, NTT 情報通信処理研究所

ない。

2.3 マニュアルの特徴を用いた省略要素同定

(1) 特定の言語表現を用いた同定

以下のA～Dに示す表現は、話し手・聞き手と文中に現れる名詞句との関係を制限する。堂坂⁶⁾は、これらの表現を用いて、省略されている名詞句が「話し手」「聞き手」のどちらに属するものを指すかを同定する。

本アルゴリズムでは、話し手「筆者」、聞き手「ユーザ」に加えて「説明対象物」を補完の対象物とし、これらの言語表現の性質を利用する。ただし、この手法は、能動文の省略主語同定には用いることができるが、受動文の省略動作主同定には用いることができない。

A: 発話行為のための言語表現を利用した省略同定
発話することによってある行為が遂行される場合、その行為を発話行為という⁷⁾。マニュアル文章における省略主語同定には行為指導型と陳述表示型を利用することができる。

例えば、行為指導型の発話行為を意味する文は、話し手が聞き手にある行為を促す行為指導の文であるということを示すマーカ（特定の動詞や様相表現）と、聞き手がなすべき行為の内容を表す命題から構成される。マニュアルの主目的は、筆者があるものの操作法や機能をユーザに伝えることであるから、筆者は話し手、ユーザは聞き手としてとらえることができる。したがって、行為指導型の動詞が使われている場合、行為の内容を表す命題中の主語は、聞き手すなわち「ユーザ」と同定することができる。

例:「～てください」(許可の行為指導を表す様相)

「～のがよい」(推薦の行為指導を表す様相)

→～(埋込み文の動詞)の主語=ユーザ

同様に、陳述表示型の発話行為を意味する文は、話し手の聞き手に対する陳述表示であることを示すマーカ（特定の動詞）と、伝達内容を表す命題から構成される。したがって、命題中の主語は、書き手「筆者」と同定することができる。

例:「以下に～を述べます。」(陳述表示の動詞)

→陳述表示の動詞の主語=筆者

B:「情報の縄張り」を利用した省略同定

ある種の文末表現は、主語がその記述内容を経験的に知りうる者であるとき、すなわち情報がその主語の縄張りにある⁸⁾ときにのみ使うことができる。例えば、「思う」、「考える」、「希望する」のような動詞が

主文の文末にあり、かつ現在形で使用されていれば、その主語は、話し手である「筆者」と同定することができる。

また、マニュアルにおける「説明対象物」は、聞き手である「ユーザ」ではなく話し手である「筆者」の縄張りにあるのが普通であり、さらに、話し手が自分自身について推論することはないことから、次の例のような様相表現を持つ文中の主語は、「説明対象物」と同定することができる。

例:「～にちがいない」

→主語=説明対象物

C: 待遇表現を利用した省略同定

尊敬表現は、主語の指すものが目的語の指すものより上位にあることを示し、謙譲表現は主語の指すものが目的語の指すものより下位にあることを示す。本アルゴリズムでは、マニュアルにおいて「筆者」は「ユーザ」を自分より上位に表現し、「説明対象物」は敬語表現にはかかわらないことに基づいて、待遇表現から省略主語を同定する。例えば、尊敬表現の「お～になる/いらっしゃる」の主語は「ユーザ」であり、謙譲表現の「お～にする」の主語は、「筆者」と同定することができる。

D: 視点を利用した省略同定

日本語の表現には、話し手がどの名詞句に視点をおいているかを示す表現がある⁹⁾。マニュアル文章において、「説明対象物」の操作法は、筆者「話し手」がユーザ「聞き手」の立場に立って説明するべきものであり、話し手の視点は、ユーザの視点に一致するべきである。したがって、話し手の視点がガ格名詞句にあることを示す様相表現(例1)を持つ文の主語は「ユーザ」、話し手の視点がヲ格名詞句にあることを示す様相表現(例2)を持つ文の主語は「説明対象物」と同定することができる。

例1:「～てやる」、「～てもらう」

→主語=ユーザ

例2:「～てくれる」

→主語=説明対象物

(2) マニュアルにおける動詞の性質を利用した同定

マニュアルという世界に限ると、「忘れる」、「確認する」、「選択する」等の思考作用に関する動詞は、「説明対象物」を動作主として取ることがない。また、「表示する」、「出力する」などの動詞は、「説明対象物」の動作を表す。したがって、これらの動詞に着目

表 1 能動化可否判定規則
Table 1 Rule table for deciding appropriate voice.

文章タイプ	動作主			
	ユーザ	説明対象物	筆者	不明“?”
操作説明文章	能動	受動	受動	受動
機能説明文章	受動	能動	受動	能動
マニュアル説明文章	受動	受動	能動	受動

して省略要素を同定することができる。なお、この手法は、能動文の省略主語同定、受動文の省略動作主同定、双方に用いることができる。

2.4 省略動作主と文章タイプに基づく能動化可否判定規則

以上の議論をもとに、文章タイプと受動文の動作主が判明しているときに適切な態を決定する条件を表1のように規則化した。表で能動とある場合は能動化可、受動とある場合は能動化不可を示す。動作主が不明の場合(“?”)の態は、デフォルト的な意味あいを持つ。

3. 文内の統語的制約から能動化できない受動文

文脈による制約とは独立に、文内の統語的制約から能動化できない受動文が存在する。いわゆる間接受動文は、その典型である。本アルゴリズムでは、間接受動文に加え以下のA~Dに示すタイプの受動文を能動化できないものとして扱う。

A: 能動と受動で相の解釈が変わる文

受動文を能動化すると、完了相の意味が弱まることがある。例えば、「データが入力されていた」と「データを入力していた」を比べると、前者はデータが入力済みという完了の意味が強いのに対し、後者はデータを入力中という継続の意味が強くなる。日本語の場合、特に継続性の動詞で動作の結果が残る性質の動詞に「～ている」がつくと、能動態と受動態ではそれぞれ継続と完了という別の相の解釈が強くなる。したがって、完了/継続のいずれかの意味を明確にする副詞がないかぎり、上記のような述語形態による受動文は、能動化不可とする。

B: 発話行為を表す動詞による受動文

「保証する」、「定義する」などの宣告命名型の動詞が主文において他の様相表現なしに受動態として使われていると、「一般的にそうになっている」と解釈される。しかし、これを能動化すると、「筆者がそうする

権利を持っていて、その場でその行為を行う」という解釈が成り立ってしまう。したがって、このような宣告命名型の発話行為を表す動詞による受動文は能動化不可とする。

C: 副助詞を持つ名詞句の解釈を変える受動文

「ハ」や「モ」などの副助詞は、意味役割を明確に示す格助詞を吸収してしまうことがある。そのため、その名詞句の意味役割の解釈は、それ単独で定めることができず、格助詞などによって意味役割が明示されている名詞句との関連によらなければならない場合がある。例えば次の例文において、ハ格名詞句(主題名詞句)を保存したまま能動化すると、「スタックがピクチャを集めている」という誤った解釈を生む。したがって、このような主題名詞句を含む受動文を能動化不可とする。

例: スタックは、コピー可能なピクチャが集められています。

→? スタックは、コピー可能なピクチャを集めています。

D: 能動化が複文中の他の一方の文の解釈に悪影響を与える受動文

ある特定の接続形態でつながれた複文中の名詞句は、その複文中の複数の述語に係っていると解釈される⁹⁾。受動文がこのような名詞句を持つ複文の中の一文であるときその動詞を能動化すると、名詞句の助詞を変えても変えなくても、他方の文の解釈に影響をもたらす。例えば、次のような例文において、主文を能動化すると従文におけるハ格名詞句が主文側にも係ると解釈される。このとき、主文側にガ格名詞句が存在しなければ、そのハ格名詞句が主語として解釈されてしまう場合がある。したがって、このような性質の特定の接続形態を含む複文における受動文を、能動化不可と判定する。

例: ポインタは元に戻らず、引き続き同じアイコンが選択されます。

→? ポインタは元に戻らず、引き続き同じアイコンを選択します。

4. 受動文能動化可否判定アルゴリズム

4.1 アルゴリズムの構成

提案するアルゴリズムの構成を図1に示す。

図1①の省略要素同定において、能動文の省略主語の同定においては、上記2.3節(1)の規則、2.3節(2)の規則、および従来の主題名詞句による補完手

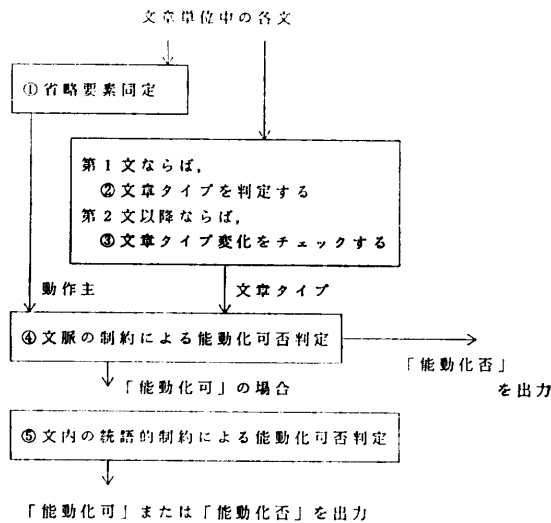


図1 能動化可否判定アルゴリズムの概要
Fig. 1 An overview of the algorithm.

法、接続形式を用いた手法⁸⁾をこの順に同定できるまで適用する。以上の手法で同定できない場合は、第1文については、省略要素不明とし、第2文以下については、前文の主語を省略要素とする。受動文の省略動作主の同定においては、2.3節(2)の規則を用いて同定を試み、同定できない場合は、省略要素不明とする。

また、能動化可否判定においては、文章タイプと同定された受動文の動作主をもとに、まず表1の能動化可否判定規則により文脈の制約において能動化可であるかを図1④で判定する。これにより能動化可と判定した場合は、さらに3章に示した文内の統語的制約による能動化可否を図1⑤で判定する。

4.2 文章タイプの判定について

2.2節で述べたように、文章タイプは文章単位を通して一定であるのが原則である。そこで、文章の第1文に対して図1②で文章タイプの判定を行う。第1文は、タイトル相当の名詞句である場合と、通常の文である場合とがある。タイトル相当の名詞句である場合、「～する方法」(操作説明文章)、「～の機能概要」(機能説明文章)などのように、文章タイプを明示する特定の表現が使われることが多いため、これらの表現を利用して判定を行う。通常の文である場合は、その動作主と態を利用して表2に示す規則により文章タイプを判定する。

上述のように、文章単位内の文章タイプ

表2 文章タイプ判定規則
Table 2 Rule table for deciding text type.

動作主 態	ユーザ	説明対象物	筆者
能動	操作説明	機能説明	マニュアル説明
受動	—	操作説明	—

は一定であるというのが原則である。しかし、現実の文章においては、文章単位内でこれが変化すると考えた方が妥当である場合がある。そこで、第2文以降の文については、図1③において文章タイプの変化をチェックする。実データの分析結果によれば、文章タイプの変化は、「機能説明文章」から「操作説明文章」の変化に限られる。また、「操作説明文章」における動作主は、「ユーザ」である。したがって、この変化は、次の規則により簡単に識別することができる。

文章タイプ変化規則：前文までの文章タイプが「機能説明文章」で処理対象文の動作主が「ユーザ」であれば、文章タイプを「操作説明文章」とする。

4.3 処理例

処理例を図2に示す。図2の文章において実線下線は、能動化可と判定される受動態述語、点線下線は、能動化不可と判定される受動態述語を示す。

5. 評価と考察

5.1 評価実験

テキストエディタの使用法、プログラミング言語の使用法についての3種類のマニュアルから無作為に抽出した計47の形式段落を文章単位とし、本アルゴリズムの精度を評価した。省略要素同定の結果を表3に、能動化可否判定の結果を表4に示す。

5.2 考察

表3(a)における成功数の約1/3が、本論文におい

入力文章
①システムは、表示する前に各メッセージのヘッダを整形します。②通常ここで、ほとんどのヘッダフィールドが消去されます。③オリジナルのヘッダは永久に保存されています。④それを見るためには、Xコマンドを使ってください。⑤オリジナルのヘッダが表示されます。

- ① 主題名詞句による省略主語判定規則により、主語 = システム「説明対象物」、文章タイプ = 「機能説明文章」。
- ② および③ 省略動作主 = 「不明(「?」)」。
- ④ 表1の規則より、能動化可と判定。文内の統語的制約からも能動化可と判定。
- ⑤ 行為指導型の発話行為を持つことから、主語 = 「ユーザ」、文章タイプ変化規則により、文章タイプ = 「操作説明文章」。
- ⑥ マニュアルの動詞を用いた規則により、主語 = 「説明対象物」、表1の規則より、能動化不可と判定。

図2 処理例
Fig. 2 An example.

表 3 省略要素同定の精度

Table 3 Results of omitted elements identification.

(a) 能動文の場合
(a) For actives.

同定手段	成功数	失敗数
特定の言語表現	28	1
マニュアルの動詞	22	0
主題名詞句	46	0
接続形式	9	0
前文の主語	53	10
主語不明	—	1
同定成功率	92.9%	

(b) 受動文の場合
(b) For passives.

同定手段	成功数	失敗数
マニュアルの動詞	20	6
動作主不明	—	38
同定成功率	31.2%	

表 4 能動化可否判定の精度

Table 4 Results of activizable passives detection.

判 定	能動化可能文	能動化不可能文
能動化可	19	5
能動化否	6	34
判定精度	76.0%	87.2%

て提案したマニュアル文章の特性を用いた同定手法によることから、能動文の省略主語同定におけるその有効性が確認された。一方、表 3 (b)からは、受動文の省略動作主同定の精度が不十分であることがわかる。表 4 によれば能動化可否の判定に失敗した文は 11 文であり、受動文の省略動作主同定率が低いことがその最大の原因となっている。受動文の省略動作主同定率が低いのは、現在のところ受動文の省略動作主の同定を 2.3 節(2)に示した方法でしか行えないことに起因している。当面これを向上させるためには、マニュアルの分野を限定して、そこで使われる動詞と動作主との結びつきをさらに分析していく必要がある。また、より本質的な解決へ向けては、マニュアルにおける文章構造の特徴(例えば、ユーザ操作を示す記述の後には、それに対するシステムの反応動作が記述されやすいなど)を考慮していく必要があると考える。

6. おわりに

マニュアル文章中の受動文の能動化可否を判定するアルゴリズムを提案した。また、そのなかで必要となる省略要素の同定について、マニュアル文章の特徴を利用する方法を提案した。提案したアルゴリズムの精度を今後さらに向上させるためには、受動文における省略動作主同定の精度向上が必要であり、それには、マニュアルにおける文章構造の特徴をも加味していく必要がある。今回提案したアルゴリズムは、「能動化できるものは、できるだけ能動化する」という条件下において、その有効性を確認することができた。しかし、本来受動態で書いた方が適切であるにもかかわらず能動態で書かれているものを取り扱うことができないなど、文脈における最適な態の選択というより広い問題に対しては、文脈における主題の流れ、視点の一貫性、動作主の有生/無生などの多くの要因をさらに検討していく必要がある。

謝辞 本研究について貴重なご助言をいただいた NTT 基礎研究所情報科学研究部吉本啓主任研究員、日頃ご指導いただいている NTT 情報通信処理研究所メッセージシステム研究部小原永主幹研究員に感謝いたします。

参 考 文 献

- 1) 林 良彦, 菊井玄一郎: 日本文推敲支援システムにおける書き換え機能, 第 4 回人工知能学会全国大会論文集, 15-6 (1990).
- 2) 高橋昭男: わかりやすいマニュアルの作成法, 日経マグロウヒル (1985).
- 3) 神尾昭雄: 談話における視点, 日本語学, Vol. 4, No. 12, pp. 10-21 (1985).
- 4) 山本祥子, 吉本 啓, 片桐恭弘, 野村浩郷: 英日翻訳における受給表現の扱いについて, 情報処理学会自然言語処理研究会, 45-2 (1984).
- 5) 久野 暉: 談話の文法, 大修館書店 (1983).
- 6) 堂坂浩二: 対話登場人物を指示する日本語ゼロ代名詞の同定, 情報処理学会談話理解モデルとその応用シンポジウム, pp. 41-52 (1989).
- 7) Searle, J.: *Speech Acts*, Cambridge University Press (1969).
- 8) 吉本 啓: 談話処理における日本語ゼロ代名詞の扱いについて, 情報処理学会自然言語処理研究会, 56-4 (1986).

(平成 2 年 5 月 30 日受付)

(平成 2 年 7 月 10 日採録)

**林 良彦 (正会員)**

1959年生。1981年早稲田大学理工学部電気工学科卒業。1983年同大学院理工学研究科博士前期課程修了。同年日本電信電話公社入社。現在、NTT 情報通信処理研究所メッセージシステム研究部主任研究員。日英機械翻訳システムの研究、日本文推敲支援システムの研究に従事。人工知能学会会員。

**千葉 裕子 (正会員)**

昭和37年生。昭和60年神戸大学教育学部卒業。昭和62年カリフォルニア大学バークレー校大学院言語学部修士課程修了。昭和63年日本電信電話(株)入社。情報通信処理研究所にて自然言語処理の研究に従事。平成2年2月退社。日本言語学会、アメリカ音響学会各会員。旧姓最上。