(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3580760号
(P3580760)
(45)発行日　平成16年10月27日(2004.10.27)　(24)登録日　平成16年7月30日(2004.7.30)

(51)Int.Cl.[7]　F I
G06F 17/21　G06F 17/21　550J

請求項の数 8　(全 11 頁)

(21)出願番号　特願2000-156765(P2000-156765)
(22)出願日　平成12年5月26日(2000.5.26)
(65)公開番号　特開2001-337945(P2001-337945A)
(43)公開日　平成13年12月7日(2001.12.7)
審査請求日　平成14年7月12日(2002.7.12)

(73)特許権者　000005049
シャープ株式会社
大阪府大阪市阿倍野区長池町２２番２２号
(74)代理人　100065248
弁理士　野河　信太郎
(72)発明者　吉見　毅彦
大阪府大阪市阿倍野区長池町２２番２２号
シャープ株式会社内

審査官　石川　正二

(56)参考文献　特開平１１－３２８１７５（ＪＰ，Ａ）
特開平０３－２７６３６７（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G06F 17/21　550

(54)【発明の名称】自動編集装置及び方法並びにこれに利用される記憶媒体

(57)【特許請求の範囲】
【請求項１】
辞書テーブル、形態素解析規則テーブル、構文解析規則テーブル及び構文木マッチング規則テーブルを記憶したテーブルメモリと、
自然言語で記述された複数の文からなる文書を入力する入力部と、
辞書テーブル及び形態素解析規則テーブルを参照し、入力された文書中の各文に対して形態素解析を行う形態素解析部と、
構文解析規則テーブルを参照し、形態素解析された各文を構文解析し、構文解析結果を構文木として出力する構文解析部と、
構文木マッチング規則テーブルを参照し、見出しの文の構文木と本文中の文の構文木がマッチするか否かを判定し、二つの構文木がマッチした場合、本文中の文の構文木の時制形態や相形態を認識する認識部と、
見出しの文の時制形態や相形態を、認識した時制形態や相形態に書き換える書き換え部とを備えたことを特徴とする自動編集装置。
【請求項２】
前記テーブルメモリは、単語補完規則テーブルを更に記憶し、
前記単語補完規則テーブルを参照し、形態素解析された各文のうち、必要な単語が省略された特有表現の文か通常表現の文かを判定し、特有表現の文ならば、省略された単語を推測して補完する単語補完部を更に備え、
前記認識部は、構文木マッチング規則テーブルを参照し、特有表現の文の構文木と通常表

現の文の構文木がマッチするか否かを判定し、二つの構文木がマッチした場合、通常表現の文の構文木の時制形態や相形態を認識し、
前記書き換え部は、必要な単語を補完した文の時制形態や相形態を、認識した時制形態や相形態に書き換えることを特徴とする請求項１に記載の自動編集装置。
【請求項３】
前記特有表現の文の構文木が文書中の表題文の構文木であり、通常表現の文の構文木が表題文に対応する文書中のいずれか１つの文の構文木であることを特徴とする請求項２記載の自動編集装置。
【請求項４】
前記特有表現の文の構文木が文書中の新聞記事の見出しの構文木であり、通常表現の構文木が新聞記事の見出しに対応した文書中のいずれか１つの文の構文木であることを特徴とする請求項２記載の自動編集装置。
【請求項５】
前記書き換え部は、特有表現の文の構文木の節の時制形態や相形態を、通常表現の文の構文木の節の時制形態や相形態に書き換えることを特徴とする請求項２記載の自動編集装置。
【請求項６】
前記単語補完部は、前記単語補完規則テーブルを参照し、形態素解析された各文のうち、ｂｅ動詞が省略された特有表現の文か通常表現の文かを判定し、特有表現の文ならば、省略されたｂｅ動詞を推測して補完することを特徴とする請求項２記載の自動編集装置。
【請求項７】
辞書テーブル、形態素解析規則テーブル、構文解析規則テーブル及び構文木マッチング規則テーブルをテーブルメモリに記憶し、
入力部を用いて、自然言語で記述された複数の文からなる文書を入力し、
形態素解析部を用いて、辞書テーブル及び形態素解析規則テーブルを参照し、入力された文書中の各文に対して形態素解析を行い、
構文解析部を用いて、構文解析規則テーブルを参照し、形態素解析された各文を構文解析し、構文解析結果を構文木として出力し、
認識部を用いて、構文木マッチング規則テーブルを参照し、見出しの文の構文木と本文中の文の構文木がマッチするか否かを判定し、二つの構文木がマッチした場合、本文中の文の構文木から時制形態や相形態を認識し、
書き換え部を用いて、見出しの文の時制形態や相形態を、認識した時制形態や相形態に書き換えることを特徴とする自動編集方法。
【請求項８】
自動編集装置に利用される記憶媒体であって、
辞書テーブル、形態素解析規則テーブル、構文解析規則テーブル及び構文木マッチング規則テーブルをテーブルメモリに記憶する機能と、
入力部を用いて、自然言語で記述された複数の文からなる文書を入力する機能と、
形態素解析部を用いて、辞書テーブル及び形態素解析規則テーブルを参照し、入力された文書中の各文に対して形態素解析を行う機能と、
構文解析部を用いて、構文解析規則テーブルを参照し、形態素解析された各文を構文解析し、構文解析結果を構文木として出力する機能と、
認識部を用いて、構文木マッチング規則テーブルを参照し、見出しの文の構文木と本文中の文の構文木がマッチするか否かを判定し、二つの構文木がマッチした場合、本文中の文の構文木から時制形態や相形態を認識する機能と、
書き換え部を用いて、見出しの文の時制形態や相形態を、認識した時制形態や相形態に書き換える機能とをコンピュータに実行させる自動編集プログラムを記憶した記憶媒体。
【発明の詳細な説明】
【０００１】
【発明が属する技術分野】

本発明は、機械翻訳システムなどの自然言語処理システムに適用され、ある自然言語文を構文解析して他の自然言語文に翻訳する前に、通常の表現形式とは異なる特有の形式をしている文を検索して通常の表現形式の文に書き換える自動編集装置及び方法並びにこれに利用される記憶媒体に関する。

【０００２】

【従来の技術】

近年、ＷＷＷ（ワールド・ワイド・ウェブ）を通じて英字新聞記事に接する機会が増えてきたことに伴い、より正確に英字新聞記事を日本語に翻訳する必要性が高まってきている。英字新聞記事は見出し（ヘッドライン）と本文から構成されるが、見出しは記事の最も重要な情報を伝える表現であるため、見出しを正確に翻訳することは他の表現の翻訳に比べてより一層重要である。

【０００３】

英字新聞記事の見出しは、できるだけ少ない文字数でできるだけ多くの情報を伝えるためや、読者の注意を引くために、通常の文の表現形式とは異なる特有の形式をしている。このため、従来の英日機械翻訳システムでは適切に翻訳できない場合が多い。その原因は主に、見出し特有の表現に対して構文解析を適切に行うことができないことにある。これは、機械翻訳システムの構文解析規則が、標準的な表現を扱うことを前提に記述されいると考えられる。

【０００４】

この問題に対し、文献「英字新聞記事見出し翻訳の自動前編集による改善」（言語処理学会第５回年次大会発表論文集、１９９９年３月、４５８～４６１頁）の記載によれば、英字新聞記事の見出しを通常の表現形式に書き換える自動前編集系を既存の機械翻訳システムに追加することによって翻訳の品質の改善を図っている。

【０００５】

上記文献（以下、先行技術と呼ぶ）によれば、例えば、見出し特有の表現のうち比較的高い頻度で見られるｂｅ動詞の省略現象に対象を絞り、ｂｅ動詞が省略されている見出しにｂｅ動詞を正しく補う方法が開示されている。

例えば、“Ｓａｌｅｓ　ｕｐ　ｓｈａｒｐｌｙ　ｉｎ　Ｊｕｎｅ”という見出しは、通常の機械翻訳システムでは適切に構文解析できない可能性が高いが、この自動前編集方法によって、例えば“Ｓａｌｅｓ　ａｒｅ　ｕｐ　ｓｈａｒｐｌｙ　ｉｎ　Ｊｕｎｅ”のように、ｂｅ動詞“ａｒｅ”を補えば、従来の機械翻訳システムでも適切な翻訳が得られるようになる。

【０００６】

【発明か解決しようとする課題】

上記の先行技術は、例えば、ｂｅ動詞を補うべき見出し（ｂｅ動詞が省略されているもの）と補うべきでない見出し（省略されていないもの）とを区別することはできる。しかし、先行技術は、ｂｅ動詞の時制や相（ａｓｐｅｃｔ）の決定を行うことはできないので、ｂｅ動詞の屈折形の生成はできない。

例えば、見出し“Ｓａｌｅｓ　ｕｐ　ｓｈａｒｐｌｙ　ｉｎ　Ｊｕｎｅ”に補うｂｅ動詞の時制（ｔｅｎｓｅ　）は、“現在時制”に固定されているため、つまり“ａｍ”、“ａｒｅ　”、“ｉｓ”のいずれかしか生成されない。本来、同じ文書中の他の文に記述されている内容から、ｂｅ動詞は、現在形の“ａｒｅ”ではなく、過去形の“ｗｅｒｅ”とすべきであるが、このような時制処理は実現されていない。

【０００７】

本発明は以上の事情を考慮してなされたものであり、例えば、自然言語の文書から必要な単語を省略した特有表現の文を検索した際に、特有表現の文に対して省略された単語を推測して補完するとともに同じ文書中の他の通常表現の文の時制形態や相形態を認識し、必要な単語を補完した文の形態を認識した時制形態や相形態に書き換えることができる自動編集装置及び方法並びにこれに利用される記憶媒体を提供する。

【０００８】

【課題を解決するための手段】
本発明は、辞書テーブル、形態素解析規則テーブル、単語補完規則テーブル、構文解析規則テーブル及び構文木マッチング規則テーブルを記憶したテーブルメモリと、自然言語で記述された複数の文からなる文書を入力する入力部と、辞書テーブル及び形態素解析規則テーブルを参照し、入力された文書中の各文に対して形態素解析を行う形態素解析部と、単語補完規則テーブルを参照し、形態素解析された各文のうち、必要な単語が省略された特有表現の文か通常表現の文かを判定し、特有表現の文ならば、省略された単語を推測して補完する単語補完部と、構文解析規則テーブルを参照し、規則テーブル形態素解析された各文を構文解析し、構文解析結果を構文木として出力する構文解析部と、構文木マッチング規則テーブルを参照し、特有表現の文の構文木と通常表現の文の構文木がマッチするか否かを判定し、二つの構文木がマッチした場合、通常表現の文の構文木から時制形態や相形態を認識する認識部と、必要な単語を補完した文の形態を認識した時制形態や相形態に書き換える書き換え部とを備えたことを特徴とする自動編集装置である。
【０００９】
本発明によれば、自然言語の文書から必要な単語を省略した特有表現の文を検索した際に、特有表現の文に対して省略された単語を推測して補完するとともに同じ文書中の他の通常表現の文の時制形態や相形態を認識し、必要な単語を補完した文の形態を認識した時制形態や相形態に書き換えることができる。よって、本発明の自動編集装置を機械翻訳装置に用いることにより、自然言語文の翻訳の品質を向上することができる。
【００１０】
前記特有表現の文の構文木が文書中の表題文の構文木であり、通常表現の文の構文木が表題文に対応する文書中のいずれか１つの文の構文木である構成にしてもよい。
【００１１】
前記特有表現の文の構文木が文書中の新聞記事の見出しの構文木であり、通常表現の構文木が新聞記事の見出しに対応した文書中のいずれか１つの文の構文木である構成にしてもよい。
【００１２】
前記書き換え部は、特有表現の文の構文木の節の形態を、通常表現の文の構文木の節の制形態や相形態に書き換える構成にしてもよい。
【００１３】
前記特有表現の文が文書中の英字新聞記事の見出しであり、英字新聞記事の見出しの省略された必要な単語がｂｅ動詞である構成にしてもよい。
【００１４】
本発明の別の観点によれば、辞書テーブル、形態素解析規則テーブル、単語補完規則テーブル、構文解析規則テーブル及び構文木マッチング規則テーブルをテーブルメモリに記憶し、入力部を用いて、自然言語で記述された複数の文からなる文書を入力し、形態素解析部を用いて、辞書テーブル及び形態素解析規則テーブルを参照し、入力された文書中の各文に対して形態素解析を行い、単語補完部を用いて、単語補完規則テーブルを参照し、形態素解析された各文のうち、必要な単語が省略された特有表現の文か通常表現の文かを判定し、特有表現の文ならば、省略された単語を推測して補完し、構文解析部を用いて、構文解析規則テーブルを参照し、形態素解析された各文を構文解析し、構文解析結果を構文木として出力し、認識部を用いて、構文木マッチング規則テーブルを参照し、特有表現の文の構文木と通常表現の文の構文木がマッチするか否かを判定し、二つの構文木がマッチした場合、通常表現の文の構文木から時制形態や相形態を認識し、書き換え部を用いて、必要な単語を補完した文の形態を認識した時制形態や相形態に書き換えることを特徴とする自動編集方法が提供される。
【００１５】
特に、英字新聞記事の見出しは、述語の時制や相や態などに関する情報が頻繁に省略される。しかし、時制や相や態などの決定に必要な情報は、英字新聞記事の本文中には明示されていることが多いので、英字新聞記事の本文中の文を参照することにより、見出しに明

示されていない時制や相や態などを決定することができる。
【００１６】
具体的には、見出しの構文木と記事本文中の各文の構文木とのマッチングを調べ、もし、見出しの構文木とマッチする文が存在する場合、その文のマッチ部分の時制や相や態などを見出しの時制や相や態などとする。
例えば、英字新聞記事の見出しを構文解析し、通常の表現形式に書き換える、ｂｅ動詞の屈折形を、人称と数に基づいて区別するだけでなく、時制や相も考慮に入れて決定することができる。
【００１７】
本発明は、このような考え方に基づくものであるが、この考えは次のように一般化できる。通常の表現形式とは異なる特有の形式をしている文を、通常の表現形式の文に書き換える際、その文に明示されていない必要な情報は、文書中の他の文で明示されている可能性があるので、必要な情報を発見することによって、文の書き換え精度を向上させることができる。
【００１８】
以下、本発明の自然言語の自動編集機能を、英字新聞記事の見出しを対象とした場合について説明するが、本発明は、英字新聞記事の見出しに限らず、他の種類の文書のタイトル（表題）及び文章や文節の表題などの編集を対象とすることもでき、さらには一般の文を対象とすることもできる。
【００１９】
【本発明の実施の形態】
以下、図に示す実施例に基づいて本発明を詳述する。なお、本発明はこれによって限定されるものではない。
【００２０】
図１は本発明の一実施例である自動編集装置の構成を示すブロック図である。図１に示すように、本発明の自動編集装置は、制御部１、入力部２、出力部３、テーブルメモリ４、プログラムメモリ５、バッファメモリ６、制御プログラムデータ及びアドレスデータを転送するバス７、記憶媒体８から構成されている。
【００２１】
制御部１は、例えば、コンピュータのＣＰＵ（中央処理装置）から構成され、プログラムメモリ５から制御プログラムを読み出し、この制御プログラムによりバス７を介して各部を制御することにより本発明の自動編集機能を実現する。入力部２は、例えば、キーボード、マウス、ペン、タブレット、スキャナ、文字認識装置などの入力装置や、通信回線と接続されている通信装置、記憶媒体読取装置などから構成され、入力部２は自然言語で記述された文書の入力、自動編集開始の指示、文書データの通信、制御プログラムのインストールなどを行う。
【００２２】
出力部３は、例えば、ＣＲＴ（陰極線管）ディスプレイ、ＬＣＤ（液晶ディスプレイ）、ＰＤ（プラズマディスプレイ）などからなる表示装置や、サーマルプリンタ、レーザプリンタなどからなる印字装置、または通信回線と接続されている通信装置で構成され、出力部３は、入力部２による入力結果、制御部１の制御により、自動編集結果や翻訳結果を表示装置に表示したり、印字装置を介して印字したり、通信装置を介して送信する。
【００２３】
テーブルメモリ４は、例えば、マスクＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系等を含めた記憶媒体から構成される
【００２４】
また、テーブルメモリ４は、単語、品詞情報を記憶した辞書テーブル４ａ、文書を形態素

解析するための形態素解析規則を記憶した形態素解析規則テーブル４ｂ、見出しのｂｅ動詞などの単語を補完するための単語補完規則を記憶した単語補完規則テーブル４ｃ、文を構文解析するための構文解析規則を記憶した構文解析規則テーブル４ｄ、構文木マッチングを行うための構文木マッチング規則を記憶した構文木マッチング規則テーブル４ｅ、類義語を記憶した類義語辞書テーブル４ｆとして機能する。

【００２５】

プログラムメモリ５は、例えば、マスクＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系等を含めた記憶媒体から構文される。

【００２６】

また、プログラムメモリ５は、形態素解析部５ａ、単語補完部５ｂ、構文解析部５ｃ、認識部５ｄ、書き換え部５ｅとして機能する各制御プログラムを記憶している。

【００２７】

バッファメモリ６は、例えば、ＲＡＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系等を含めた記憶媒体から構文される。

【００２８】

また、バッファメモリ６は、入力部１より入力された文書を記憶する文書バッファ６ａ、形態素解析結果を記憶する形態素解析結果バッファ６ｂ、単語補完結果を記憶する単語補完結果バッファ６ｃ、構文解析結果を記憶する構文解析結果バッファ６ｄ、書き換え結果を記憶する書き換え結果バッファ６ｅとして機能する領域に備えている。

書き換え結果バッファ６ｅに記憶されている内容は、バス７を介して出力部３に出力される。

【００２９】

また、図１において、形態素解析部５ａは、文書バッファ６ａに記憶されている各文に対して、辞書テーブル４ａと形態素解析規則テーブル４ｂとを参照しながら形態素解析を行い、文中の各語について品詞などの形態素・語彙属性を出力する。その形態素解析結果をバッファメモリ６中の形態素解析結果バッファ６ｂに記憶する。

【００３０】

単語補完部５ｂは、形態素解析結果バッファ６ｂに記憶されている形態素解析結果に対して、単語補完規則テーブル４ｃを参照しながら必要な単語の補完を行い、補完した単語を単語補完結果バッファ６ｃに記憶する。

例えば、単語補完部５ｂは、英字新聞記事の見出しのｂｅ動詞を補完することができるが、このｂｅ動詞補完処理は、入力された文書が英字新聞記事の見出しである場合にのみ必要な処理である。

【００３１】

構文解析部５ｃは、形態素解析結果バッファ６ｂや単語補完結果バッファ６ｃに記憶されている形態素、語彙属性列に対して、構文解析規則テーブル４ｄを参照しながら構文解析を行い、構文解析結果から得られた構文木を構文解析結果バッファ６ｄに記憶する。

【００３２】

認識部５ｄは、構文解析結果バッファ６ｄに記憶されている各構文木に対して、構文木マッチング規則テーブル４ｅと類義語辞書テーブル４ｆを参照しながら、二つの構文木（特有表現の文の構文木と通常表現の文の構文木）がマッチするか否かを判定し、二つの構文木がマッチした場合、単語を補完しない通常表現の文の構文木から時制形態や相形態を認識する。

書き換え部５ｅは、二つの構文木のマッチに成功した場合、特有表現の文の構文木の節の

形態を、通常表現の文の構文木の節の時制形態や相形態に書き換え、その書き換え結果を書き換え結果バッファ６ｅに記憶する。つまり、書き換え部５ｅは、必要な単語を補完した文の形態を、認識部５ｄで認識した時制形態や相形態に書き換える。

【００３３】

また、本発明の自動編集機能を実現するために、辞書テーブル４ａ、形態素解析規則テーブル４ｂ、単語補完規則テーブル４ｃ、構文解析規則テーブル４ｄ及び構文木マッチング規則テーブル４ｅをテーブルメモリ４に記憶する機能と、入力部２を用いて、自然言語で記述された複数の文からなる文書を入力する機能と、形態素解析部５ａを用いて、辞書テーブル４ａ及び形態素解析規則テーブル４ｂを参照し、入力された文書中の各文に対して形態素解析を行う機能と、単語補完部５ｂを用いて、単語補完規則テーブル４ｃを参照し、形態素解析された各文のうち、必要な単語が省略された特有表現の文か通常表現の文かを判定し、特有表現の文ならば、省略された単語を推測して補完する機能と、構文解析部５ｃを用いて、構文解析規則テーブル４ｄを参照し、形態素解析された各文を構文解析し、構文解析結果を構文木として出力する機能と、認識部５ｄを用いて、構文木マッチング規則テーブル４ｅを参照し、特有表現の文の構文木と通常表現の文の構文木がマッチするか否かを判定し、二つの構文木がマッチした場合、通常表現の文の構文木から時制形態や相形態を認識する機能と、書き換え部５ｅを用いて、必要な単語を補完した文の形態を認識した時制形態や相形態に書き換える機能とをコンピュータに実行させる自動編集プログラムを記憶した記憶媒体８を利用してもよい。

【００３４】

記憶媒体８は、例えば、マスクＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、フラッシュＲＯＭ等による半導体メモリ、あるいは磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやＣＤ－ＲＯＭ／ＭＯ／ＭＤ／ＤＶＤ等の光ディスクのディスク系、ＩＣカード（メモリカードも含む）／光カード等のカード系等を含めた本体と分離可能なメディアで構成した固定的にプログラムを担持する記憶媒体を示し、記憶媒体８に本発明の自然言語自動編集プログラムを記憶し、入力部２の記憶媒体読取装置を介してバッファメモリ６の予備領域に自動編集プログラムをインストールすることにより本発明の自動編集機能を実現してもよい。

【００３５】

また、この記憶媒体８は、本自動編集装置がインターネットを含めた外部の通信ネットワークとの接続可能な通信装置を備えている場合には、その通信装置を介して通信ネットワークからプログラムをダウンロードするように流動的にプログラムを担持する媒体であってもよい。なお、このように通信ネットワークからプログラムをダウンロードする場合には、そのダウンロード用プログラムは予め本体装置に格納しておくか、あるいは別な記憶媒体からインストールされるものであってもよい。なお、記憶媒体８に格納されている内容としてはプログラムに限定されず、データであってもよい。

【００３６】

図２は本実施例の自動編集装置の処理手順を示すフローチャートである。図２に用いて、本発明の自動編集装置の処理手順を、英字新聞記事の見出し書き換え処理として説明する。

Ｓｔｅｐ１：形態素解析部５ａは、文書バッファ６ａに記憶されている英字新聞記事の見出し（ヘッドライン）に対して、辞書テーブル４ａと形態素解析規則テーブル４ｂを参照しながら形態素解析を行う。その形態素解析結果を形態素解析結果バッファ６ｂに記憶する。

【００３７】

この形態素解析は、非常によく知られている一般的な技術であり、例えば、文献「自然言語処理」（長尾眞　岩波書店１９９７）などに解説があるので、説明は略す。

【００３８】

Ｓｔｅｐ２：単語補完部５ｂは、単語補完規則テーブル４ｃを参照しながら、形態素解析結果に対して、見出しにｂｅ動詞の補完が必要な場合に先行技術に基づいてｂｅ動詞を補

完する。このｂｅ動詞補完処理が行われた見出しは、単語補完結果バッファ６ｃに記憶される。

【００３９】

例えば、見出し“Ｓａｌｅｓ ｕｐ ｓｈａｒｐｌｙ ｉｎ Ｊｕｎｅ”に対して処理を行うと、“Ｓａｌｅｓ ａｒｅ ｕｐ ｓｈａｒｐｌｙ ｉｎ Ｊｕｎｅ”を形態素解析して得られる結果と同じ結果が単語補完結果バッファ６ｃに記憶される。

また、“Ｇｏｖｅｒｎｍｅｎｔ ａｐｐｒｏｖｅｓ‘ｂｒｉｄｇｅ ｂａｎｋ’ｓｃｈｅｍｅ”という見出しを処理した場合、この見出しに対しては、ｂｅ動詞は補完されないので、単語補完結果バッファ６ｃには形態素解析結果バッファ６ｂの内容と同じ内容が記憶される。

【００４０】

Ｓｔｅｐ３：構文解析部５ｃは、構文解析規則テーブル４ｄを参照しながら、単語補完結果に対して構文解析を行い、その構文解析結果（新聞記事の見出しの構文木）を構文解析結果バッファ６ｄに記憶する。構文解析処理も、形態素解析処理と同じく公知の技術であるので、説明は省略する。

Ｓｔｅｐ４：制御部１は、現在、処理中の文が新聞記事の本文中の何番目の文であるかを示すカウンタの数値ｉを１にセットする。

【００４１】

Ｓｔｅｐ５：形態素解析部５ａは、新聞記事の本文中の第ｉ番目の文に対して辞書テーブル４ａと形態素解析規則テーブル４ｂを参照しながら形態素解析を行い、その形態素解析結果を形態素解析結果バッファ６ｂに記憶する。続いて、構文解析部５ｃは、形態素解析結果に対して構文解析規則テーブル４ｄを参照しながら、構文解析を行い、その構文解析結果（新聞記事の本文中の第ｉ番目の文の構文木）を構文解析結果バッファ６ｄに記憶する。

【００４２】

Ｓｔｅｐ６：認識部５ｄは、構文木マッチング規則テーブル４ｅと類義語辞書テーブル４ｆを参照しながら、構文解析結果に対して、見出しの構文木と、新聞記事の本文中の第ｉ番目の文の構文木とがマッチするかどうかを調べる。マッチすれば、Ｓｔｅｐ７の処理へ移行し、マッチしなければ、Ｓｔｅｐ８の処理へ移行する。

Ｓｔｅｐ７：認識部５ｄは、第ｉ文のマッチした構文木の時制を、見出しのマッチした構文木の時制に決定し、書き換え部５ｅは、補完したｂｅ動詞に対し、時制処理が行って処理を終える。

【００４３】

Ｓｔｅｐ８：処理の終了条件を調べ、もし、終了条件が成り立てば処理を終える。もし、終了条件が成り立たなければ、Ｓｔｅｐ９へ移行する。

Ｓｔｅｐ９：もし、終了条件が成り立たなければ、文カウンターｉを１増やしてＳｔｅｐ５に戻る。終了するかどうかは、カウンタの数値ｉがある一定値ｎを越えるかどうかで判断する。

【００４４】

ｎとしては、処理対象の記事の全文数をとってもよいし、あるいは記事の第一段落に含まれる文数をとってもよいし、あるいは、見出しが記事の第一文とマッチする可能性は、他の文とマッチする可能性よりもかなり高いので、ｎ＝１としてもよい。

【００４５】

ここで、Ｓｔｅｐ６の処理について、下記に示す英字新聞記事の自動編集例を挙げながら詳細に説明する。なお、Ｈは見出しを示し、Ｓｉは記事本文の第ｉ文を示すものとする。

Ｈ Ｇｏｖｅｒｎｍｅｎｔ ａｐｐｒｏｖｅｓ‘ｂｒｉｄｇｅ ｂａｎｋ’ｓｃｈｅｍｅ

Ｓ１ Ｔｈｅ ｇｏｖｅｒｎｍｅｎｔ ｏｎ Ｔｈｕｒｓｄａｙ ａｐｐｒｏｖｅｄ ａ “ｂｒｉｄｇｅ ｂａｎｋ ”ｐｌａｎ ｔｏ ｔａｋｅ ｏｖｅｒ ｂａｎｋｓ ｔｈａｔ ｆａｉｌ ａｎｄ ｅｘｔｅｎｄ ｌｏａｎｓ ｔｏ ｓｏｕｎｄ ｂｏｒｒｏ

ｗｅｒｓ．
Ｓ２　Ｔｈｅ　ｐｌａｎ　ｗａｓ　ｂａｓｅｄ　ｏｎ　ａ　ｄｒａｆｔ　ａｐｐｏｖｅｄ　ａｎｄ　ａｎｎｏｕｎｃｅｄ　ｂｙ　ｒｕｌｉｎｇ　Ｌｉｂｅｒａｌ　Ｄｅｍｏｃｒａｔｉｃ　Ｐａｒｔｙ　ｅａｒｌｉｅｒ　ｉｎ　ｔｈｅ　ｄａｙ．

【００４６】

Ｓｔｅｐ５までの処理において、上記の英字新聞記事の見出しの構文解析と第一文の構文解析が終了しており、図３に示す英字新聞記事の構文木が得られているものとする。

図３は本実施例の構文解析結果から取得した新聞記事の構文木の構造例を示す図である。図３（ａ）は見出しの構文木の構造例を示す。図３（ｂ）は第一文の構文木の構造例を示す。図３に示すように、構文木の枝には節点とその子節点との関係を示すラベルが付与されている。

【００４７】

例えば、ラベル“ＡＧＴ”は、子節点“ｇｏｖｅｒｎｍｅｎｔ”が、節点“ａｐｐｏｒｏｖｅ”の行為者格であることを意味し、“ＯＢＪ”、“ＴＩＭＥ”、及び“ＧＯＡＬ”は、それぞれ目的格、時間格、目標格を意味する。

図３（ｂ）において、第一文の構文木の“ｔａｋｅ　ｏｖｅｒ”を根節点とする部分構文木の構造は省略する。

【００４８】

ここで、構文木の包含関係を次のように定める。

定義：構文木Ｘが、構文木Ｙに含まれるとは、以下の関係を満たす場合をいう。Ｘの根節点（親節点を持たない節点）Ｒ自体あるいはＲの類義語がＹ上に存在する。このとき、Ｒ自体あるいはＲの類義語を根節点とするＹの部分構文木をＺとすると、ＸとＺが次の条件１または条件２を満たす。

【００４９】

条件１：Ｘの根節点Ｒが終端節点（子節点を持たない節点）ならば、Ｒ自体あるいはＲの類義語がＺの根節点である。

条件２：Ｘの根節点Ｒが非終端節点ならば、Ｒ自体あるいはＲの類義語がＺの根節点であり、かつ、Ｒのすべての子節点Ｎ１，Ｎ２，……，Ｎｎ　について、ＲとＮｉ（１≦ｉ≦ｎ）との関係がＺにおいて成立する。さらに、この関係を満たすＺの部分構文木をＺｉとするとき、Ｎｉを根節点とするＸの部分構文木をＸｉと、Ｚｉとの間で条件１または２が成り立つ。この定義は、構文木マッチング規則テーブル４ｅに記憶されている。（なお、上の説明では条件を自然言語で記載しているが、実際には自動編集装置が一義に認識可能なデータ形式で符号化して記憶されていることは言うまでもない。）

【００５０】

Ｓｔｅｐ６の構文木のマッチング処理では、新聞記事の見出しの構文木と、新聞記事の本文中の第ｉ番目の文の構文木の間で、上記の定義に定めた包含関係が成り立つ場合、二つの構文木がマッチするものとみなす。なお、格節点に対応する語句の類義語は、類義語辞書テーブル４ｆによって求めることができるものとする。

【００５１】

今、図３の見出しＨの構文木をＸとし、記事第一文Ｓ１　の構文木をＹとする。このとき、Ｘの根節点“ａｐｐｒｏｖｅ”はＹの根節点として存在するので、Ｙそのものが部分構文木Ｚとなる。

ＸとＺについて条件１または２が成り立つかを調べる。明らかに、Ｘの根節点とＺの根節点は一致する。Ｘにおける根節点“ａｐｐｒｏｖｅ”とその子節点との関係“ＡＧＴ”、“ＯＢＪ”は、Ｚにおいても成り立っている。

【００５２】

従って、“ｇｏｖｅｒｎｍｅｎｔ”を根節点とするＸの部分構文木$Ｘ_1$と、同じく“ｇｏｖｅｒｎｍｅｎｔ”を根節点とするＺの部分構文木$Ｚ_1$　との間で、条件１または条件２が成り立つかを調べると、条件１が成り立つことがわかる。

【００５３】

同様に、“ｓｃｈｅｍｅ”を根節点とするＸの部分構文木X_2と、“ｐｌａｎ”を根節点とするＺの部分構文木Z_2との間で条件が成り立つかどうかを調べるが、類義語辞書テーブル４ｆに“ｓｃｈｅｍｅ”と“ｐｌａｎ”が類義語関係にあることが記述されているものとする。これら二つの部分構文木X_2とZ_2も条件を満たすことがわかる。

【００５４】

以上の処理により、見出しＨの構文木が第一文Ｓ１の構文木に含まれることになり、図２のＳｔｅｐ６からＳｔｅｐ７へ移行し、第一文の構文木のマッチ部分の時制を見出しのマッチした構文木の時制に決定して書き換えると、見出しは、“Ｇｏｖｅｒｎｍｅｎｔ　ａｐｐｒｏｖｅｄ‘ｂｒｉｄｇｅ　ｂａｎｋ’ｓｃｈｅｍｅ”と書き換えられる。

【００５５】

新聞記事の見出しでは、過去の事柄も現在形で表現されることが多いため、“ａｐｐｒｏｖｅ”の時制をそのまま現在と解釈することは正しくない。この問題に対して、本発明によれば、見出しでは、通常明示されていない時制情報を新聞記事の本文中の文から得ることが可能となり、時制を正しく解釈することができる。

【００５６】

【発明の効果】

本発明によれば、自然言語の文書から必要な単語を省略した特有表現の文を検索した際に、特有表現の文に対して省略された単語を推測して補完するとともに同じ文書中の他の通常表現の文の時制形態や相形態を認識し、必要な単語を補完した文の形態を認識した時制形態や相形態に書き換えることができる。よって、本発明の自動編集装置を機械翻訳装置に用いることにより、自然言語文の翻訳の品質を向上することができる。

【図面の簡単な説明】

【図１】本発明の一実施例である自動編集装置の構成を示すブロック図である。

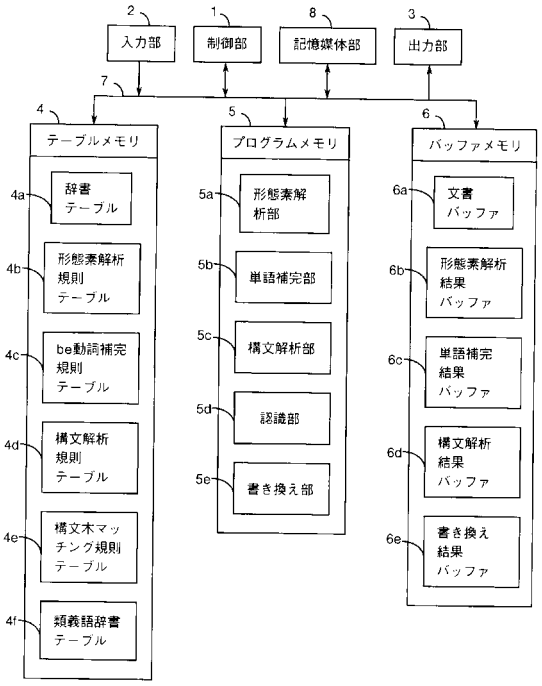
【図２】本実施例の自動編集装置の処理手順を示すフローチャートである。

【図３】本実施例の構文解析結果から取得した新聞記事の構文木の一例を示す図である。

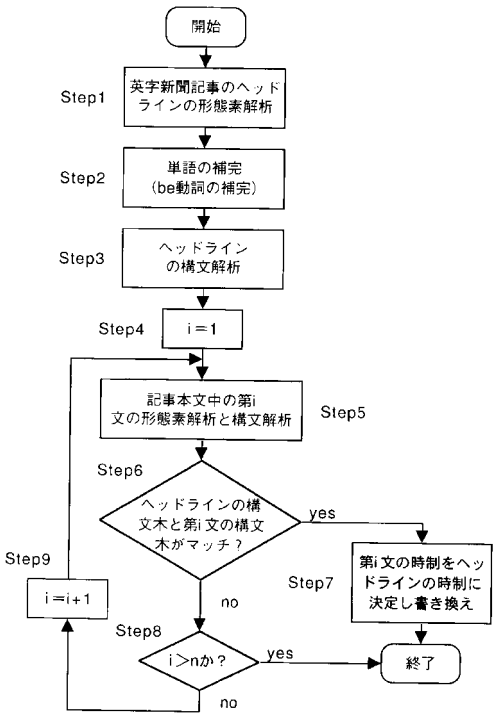
【符号の説明】

１　制御部
２　入力部
３　出力部
４　テーブルメモリ
４ａ　辞書テーブル
４ｂ　形態素解析規則テーブル
４ｃ　単語補完規則テーブル
４ｅ　構文木マッチング規則テーブル
４ｄ　類義語辞書テーブル
５　プログラムメモリ
５ａ　形態素解析部
５ｂ　単語補完部
５ｃ　構文解析部
５ｄ　認識部
５ｅ　書き換え部
６　バッファメモリ
６ａ　文書バッファ
６ｂ　形態素解析結果バッファ
６ｃ　単語補完結果バッファ
６ｄ　構文解析結果バッファ
６ｅ　書き換え結果バッファ
７　バスライン
８　記憶媒体

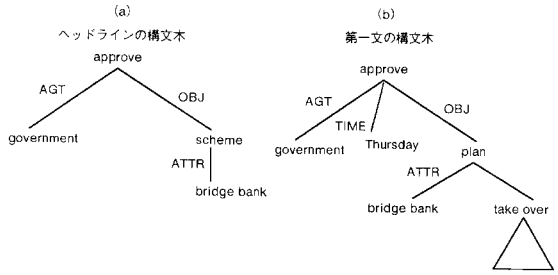
【図１】


2
入力部
1
制御部
8
記憶媒体部
3
出力部
7
4
テーブルメモリ
4a
辞書
テーブル
4b
形態素解析
規則
テーブル
4c
be動詞補完
規則
テーブル
4d
構文解析
規則
テーブル
4e
構文木マッ
チング規則
テーブル
4f
類義語辞書
テーブル
5
プログラムメモリ
5a
形態素解
析部
5b
単語補完部
5c
構文解析部
5d
認識部
5e
書き換え部
6
バッファメモリ
6a
文書
バッファ
6b
形態素解析
結果
バッファ
6c
単語補完
結果
バッファ
6d
構文解析
結果
バッファ
6e
書き換え
結果
バッファ

【図２】


開始
Step1
英字新聞記事のヘッド
ラインの形態素解析
Step2
単語の補完
(be動詞の補完)
Step3
ヘッドライン
の構文解析
Step4
i＝1
記事本文中の第i
文の形態素解析と構文解析
Step5
Step6
ヘッドラインの構
文木と第i文の構文
木がマッチ？
yes
Step7
第i文の時制をヘッ
ドラインの時制に
決定し書き換え
no
Step8
i＞nか？
yes
終了
no
Step9
i＝i＋1

【図３】


(a)
ヘッドラインの構文木
approve
AGT
OBJ
government
scheme
ATTR
bridge bank
(b)
第一文の構文木
approve
AGT
TIME
OBJ
government
Thursday
plan
ATTR
bridge bank
take over