

(51)Int.Cl.⁷
G06F 17/28

識別記号

F I
G06F 17/28

テ-マコード (参考)
P 5B091

審査請求 有 請求項の数 6 O L (全11頁)

(21)出願番号 特願2001 - 205889 (P 2001 - 205889)

(22)出願日 平成13年 7月 6日 (2001.7.6)

特許法第30条第 1 項適用申請有り 平成13年 3月30日
言語処理学会発行の「言語処理学会第 7 回年次大会ワー
クショップ論文集」に発表

(71)出願人 301022471
独立行政法人通信総合研究所
東京都小金井市貫井北町 4 - 2 - 1

(72)発明者 村田 真樹
東京都小金井市貫井北町 4 - 2 - 1 独立
行政法人通信総合研究所内

(72)発明者 井佐原 均
東京都小金井市貫井北町 4 - 2 - 1 独立
行政法人通信総合研究所内

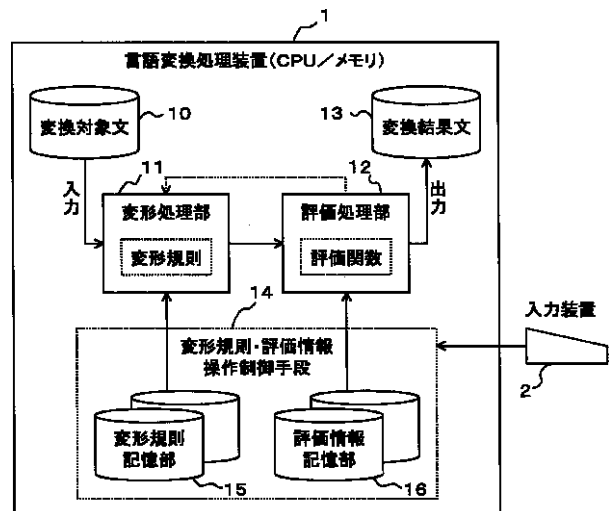
(74)代理人 100097836
弁理士 福井 國敞 (外 2 名)
F タ-ム (参考) 5B091 AA15 CA21 CC03 CC15 EA00

(54)【発明の名称】言語変換処理統一システム

(57)【要約】

【課題】 各種の言い換えを统一的に扱うことができるシステムを提供し、種々の言語変換処理システムの開発を容易に行うことができるようにする。

【解決手段】 変形処理部 1 1 は、変換対象文 1 0 を入力すると、変形規則記憶部 1 5 中の変形規則を用いて多くの変換の候補を生成する。評価処理部 1 2 は、生成された変換の候補について、文字列を変形した結果が目的とするふさわしい変換であるかどうかを評価するための尺度である評価情報を用いて評価し、評価結果のよい表現の文字列を選択する。その評価の高い文字列を変換結果文 1 3 として出力する。変換規則および評価情報を交換または選択するだけで、質問応答、文内圧縮 (要約)、難解文変換 (平易文生成)、推敲、個人文体変換といった各種の言語変換処理に適用させたシステムを構築することができる。



【特許請求の範囲】

【請求項 1】 ある自然言語で記述された文字列を、同一の自然言語で記述された他の表現による文字列に変換するシステムであって、前記自然言語の文字列に関する変形の規則を記憶する変形規則記憶手段と、文字列を変形した結果が目的とするふさわしい変換であるかどうかを評価するための尺度を与える評価関数または評価規則を記憶する評価情報記憶手段と、自然言語で記述された変換対象の文字列を入力する入力手段と、前記入力された文字列を前記変形規則記憶手段に記憶された変形の規則を用いて変形し、変換の候補を生成する変形処理手段と、前記変形処理手段によって生成された変換の候補を、前記評価情報記憶手段に記憶された評価関数または評価規則を用いて評価し、評価結果のよい表現を選択する評価処理手段と、前記変形処理手段による変形および前記評価処理手段による評価結果から、評価の高い文字列の変換結果を出力する出力手段とを備え、かつ、前記変形処理手段が用いる変形規則および前記評価処理手段が用いる評価関数または評価規則は、変換の目的に応じて交換または選択可能に構成されていることを特徴とする言語変換処理統一システム。

【請求項 2】 前記変換対象となる文字列は、質問応答システムにおける質問文と質問文に対する回答を含む文の候補の文字列であり、前記変形規則は、第 1 の文字列を同義の第 2 の文字列に言い換えるパターンの規則であり、前記評価関数または評価規則による評価の尺度は、質問文と回答を含む文の候補との類似度の大小を含み、類似度の大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【請求項 3】 前記文字列の変換は、入力した文字列から冗長な表現を修正して短くした文字列を生成する文内圧縮システムのための変換であり、前記変形規則は、第 1 の文字列を第 2 の文字列に言い換えるパターンの規則であり、前記評価関数または評価規則による評価の尺度は、変形後の文字列の長短を含み、変形後の文字列が短いものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【請求項 4】 前記文字列の変換は、入力した文字列の表現を推敲する推敲システムのための変換であり、前記変形規則は、第 1 の文字列を同義の第 2 の文字列に言い換えるパターンの規則であり、前記評価関数または評価規則による評価の尺度は、大量の用例に関する言語データ中に現れる出現頻度または出現確率を含み、出現頻度または出現確率が大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【請求項 5】 前記文字列の変換は、入力した難解文の文字列を平易文に変換する難解文変換システムのための変換であり、前記変換規則は、第 1 の文字列を同義の第

2 の文字列に言い換えるパターンの規則であり、前記評価関数または評価規則による評価の尺度は、平易な文章集合からなる大量の用例における出現頻度または出現確率を含み、出現頻度または出現確率が大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【請求項 6】 前記文字列の変換は、入力した文字列の表現をある特定の個人の文体に変換する個人文体変換システムのための変換であり、前記変換規則は、第 1 の文字列を同義の第 2 の文字列に言い換えるパターンの規則であり、前記評価関数または評価規則による評価の尺度は、前記特定の個人の文章集合からなる大量の用例における出現頻度または出現確率を含み、出現頻度または出現確率が大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ある自然言語で記述された文または文章などの文字列を、同一の自然言語で記述された他の表現による文字列に変換するシステムであって、特に、コンピュータによる自然言語処理において多種多様な言い換えを統一的に扱うことができるようにした言語変換処理統一システムに関するものである。

【0002】

【従来の技術】自然言語で記述された文または文章に関する表現の変換処理として典型的なものは、機械翻訳である。機械翻訳では、ある国の自然言語で記述された文または文章を他の国の自然言語で記述された文または文章に変換する。

【0003】機械翻訳が他の国の言語に変換するのに対し、同一の自然言語間での文または文章の変換処理を行うシステムも用いられるようになってきている。例えば、要約文を自動生成したり、文章を推敲したりするシステムである。

【0004】一般に同一自然言語間での文の変換処理では、変換前の語・句・文などのパターンと変換後の語・句・文などのパターンとの対からなる変換規則を大量に用意し、いわゆるパターン・マッチングによって入力文中に現れる変換前のパターンを探し出し、該当するパターンがあれば、それを変換後の語・句・文などのパターンに置き換える処理を行っている。

【0005】

【発明が解決しようとする課題】従来の同一自然言語間での文または文章の変換処理では、一般に変換規則による一律な変換を行っており、変換結果の良し悪しについての評価は行われていなかった。また、平易文生成、要約文生成、文章の推敲といった変換の目的に応じて、それぞれ個別に独自のシステムを構築する必要があり、あ

るシステムで使用する変換の処理ロジック（処理プログラム）を、他のシステムでも利用したり、あるシステムで使用する変換のための規則を他のシステムでも利用したりするといったことはできなかった。

【0006】そのため、複数の言語変換処理システムを開発する場合に、言語変換処理システムの種類に比例して開発コストが大きくなるという問題があった。また、言語変換処理システムの利用者（アプリケーションプログラムを含む）も、それぞれ個々のシステムごとにユーザ・インタフェースが異なるため、各システム独自のインタフェースに従ってシステムを利用しなければならず、使用方法の習得に時間がかかるという問題があった。

【0007】本発明は上記問題点の解決を図り、各種の言い換えを統一的に扱うことができるシステムを提供し、複数の言語変換処理システムを開発する場合の開発コストを低減し、また、複数の言語変換処理システムを統一されたインタフェースで利用可能にすることを目的とする。

【0008】

【課題を解決するための手段】本発明は、上記課題を解決するため、ある自然言語で記述された文字列を、同一の自然言語で記述された他の表現による文字列に変換するシステムにおいて、主要なモジュールとして変形処理手段と評価処理手段とを持ち、さらにこれらに付随する入力手段および出力手段を持つ。また、自然言語の文字列に関する変形の規則を記憶する変形規則記憶手段と、文字列を変形した結果が目的とするふさわしい変換であるかどうかを評価するための尺度を与える評価関数または評価規則を記憶する評価情報記憶手段とを持ち、これらを変換の目的に応じて交換できるようにする。または、変形規則記憶手段および評価情報記憶手段に、変形規則および評価関数等を複数種類用意し、変換の目的に応じて選択できるようにする。

【0009】入力手段によって、自然言語で記述された変換対象の文字列を入力すると、変形処理手段は、入力された文字列を変形規則記憶手段に記憶された変形の規則を用いて変形し、変換の候補を生成する。評価処理手段は、変形処理手段によって生成された変換の候補を、評価情報記憶手段に記憶された評価関数または評価規則を用いて評価し、評価結果のよい表現を選択する。出力手段は、変形処理手段による変形および評価処理手段による評価結果から、評価の高い文字列の変換結果を出力する。

【0010】ここで、変形処理手段が用いる変形規則および評価処理手段が用いる評価関数または評価規則は、変換の目的に応じて交換または選択可能に構成されているため、これらの交換または選択によって、多種多様な言い換えに対応できるようになっている。

【0011】自然言語による質問に対して、知識データ

ベースを検索して回答を自然言語で応える質問応答システムに対応する場合、入力手段は、質問文と質問文に対する回答を含む文の候補の文字列を入力する。変形規則としては、第1の文字列を同義の第2の文字列に言い換えるパターンの規則を用い、評価関数または評価規則による評価の尺度としては、質問文と回答を含む文の候補との類似度の大小であって、類似度の大きいものに高い評価を与えるものを用いる。出力手段は、類似度が大きくなるように変換された質問文と質問文に対する回答を含む文の候補の文字列とを出力し、これの照合によって回答を得られるようにする。

【0012】入力した自然言語の文字列から、冗長な表現を短くした文字列を生成する文内圧縮システムに対応する場合、変形規則としては、第1の文字列を第2の文字列に言い換えるパターンの規則を用いる。また、評価関数または評価規則による評価の尺度としては、変形後の文字列の長短であって、変形後の文字列が短いものに高い評価を与えるものを用いる。

【0013】入力した自然言語の文字列の表現を推敲する推敲システムに対応する場合、変形規則としては、第1の文字列を同義の第2の文字列に言い換えるパターンの規則を用いる。また、評価関数または評価規則による評価の尺度としては、大量の用例に関する言語データ中に現れる出現頻度または出現確率であって、出現頻度または出現確率が大きいものに高い評価を与えるものを用いる。

【0014】入力した自然言語の文字列の表現を難解文から平易文に変換する難解文変換システムに対応する場合、変換規則としては、第1の文字列を同義の第2の文字列に言い換えるパターンの規則を用いる。評価関数または評価規則による評価の尺度としては、平易な文章集合からなる大量の用例における出現頻度または出現確率であって、出現頻度または出現確率が大きいものに高い評価を与えるものを用いる。

【0015】入力した自然言語の文字列の表現をある特定の個人の文体に変換する個人文体変換システムに対応する場合、変換規則としては、第1の文字列を同義の第2の文字列に言い換えるパターンの規則を用いる。評価関数または評価規則による評価の尺度としては、特定の個人の文章集合からなる大量の用例における出現頻度または出現確率であって、出現頻度または出現確率が大きいものに高い評価を与えるものを用いる。

【0016】

【発明の実施の形態】図1は、本発明のシステム構成例を示す。図中、1はCPUおよびメモリなどからなる言語変換処理装置、2はキーボードその他の入力装置を表す。

【0017】変換対象文10は、本システムにおける入力となる自然言語文である。以下、特に断らないが変換対象文10は必ずしも一文に限られるわけではなく、文

章または句、節のようなものであってもよい。変換結果文 1 3 は、本システムの出力であって、変換対象文 1 0 を同一の種類の自然言語で言い換えたものである。

【 0 0 1 8 】言語変換処理装置 1 のモジュールは、基本的に変形処理部 1 1 と評価処理部 1 2 とから構成される。変形処理部 1 1 は、変形規則記憶部 1 5 に格納されている変形規則を用いて、変換の候補を獲得するモジュールである。評価処理部 1 2 は、変換の候補のよさを、あらかじめ評価情報記憶部 1 6 に記憶されている評価の尺度（評価関数など）によって評価し、最もふさわしい変換の候補を選択するモジュールである。

【 0 0 1 9 】変換対象文 1 0 が入力されると、変形処理部 1 1 は、変形規則を用いて変換の候補を挙げ、評価処理部 1 2 は、変形の妥当性をチェックして、最も妥当であると判断されたものを選択して、その結果を変換結果文 1 3 として出力する。

【 0 0 2 0 】変形規則記憶部 1 5 に記憶する変形規則は、人手によってあらかじめ作成された規則であってもよいし、コンピュータによって大量の言語データから自動獲得したものでもよい。変形規則および変換の候補を評価する評価関数（評価尺度）の評価情報は、扱う問題ごとに適正なものを用意する必要がある。変形規則・評価情報操作制御手段 1 4 は、利用者、システム管理者またはシステム開発者によって入力装置 2 から指示された変形規則および評価情報を、変形規則記憶部 1 5 および評価情報記憶部 1 6 に設定する制御を行う手段である。

【 0 0 2 1 】本システムで扱う問題に応じて変形規則および評価情報を適正なものにする方法として、変形規則・評価情報操作制御手段 1 4 は、例えば次のような方法を取り得る。

【 0 0 2 2 】（ 1 ）変形規則記憶部 1 5 および評価情報記憶部 1 6 を、それぞれあらかじめシステムで一意的な固定のファイル名を持つファイルで実現し、問題に応じてそれらのファイルに格納する変形規則および評価情報を書き換える。

【 0 0 2 3 】（ 2 ）変形規則記憶部 1 5 および評価情報記憶部 1 6 を、それぞれ問題に応じた複数のファイルで構成し、本システムの使用時または環境設定時に、それらのファイル中で使用する変換規則、評価情報が格納されたファイルのファイル名を入力装置 2 から入力して、変形処理部 1 1 および評価処理部 1 2 にそれぞれ通知する。

【 0 0 2 4 】（ 3 ）変形規則記憶部 1 5 および評価情報記憶部 1 6 に記憶させる各変形規則および各評価情報に、問題に応じた識別子を付与しておき、本システムの使用時または環境設定時に、扱う問題に応じた識別子を入力装置 2 から入力して、変形処理部 1 1 および評価処理部 1 2 にそれぞれ通知し、変形処理部 1 1 および評価処理部 1 2 は、入力された識別子が付与された変形規則および評価情報を使用する。

【 0 0 2 5 】評価情報は、評価のための数値情報であってもよいし、関数群もしくはサブルーチン群などによる手続き的なものであってもよい。また、評価方法を記述した規則（ルール）であってもよい。これらの組み合わせで実現することも可能である。評価処理部 1 2 で用いる評価の尺度の例としては、以下のようなものが考えられる。

【 0 0 2 6 】（ 1 ）類似度

例えば、A と B の類似度を調べたいとする。このとき、変形処理部 1 1 が使用する変形規則がすべて同義性を満足するものであるとする。この場合、A と B の類似度が大きくなるように、変形規則で A、B を変形し、A、B をよく似た状態にしてから類似度を求める。こうすることにより、意味が同じなのに異なる表現で記述されているような場合でも正しく類似度を計算することができる。なお、類似度の値は、A、B をそれぞれ構文解析し、一致する単語数、文節数、係り受け距離（構文木における二つの文節の間の枝の数）、文節距離などを考慮して定めることができる。

【 0 0 2 7 】（ 2 ）長さ

例えば、要約の一つの分野の文圧縮のように、なるべく意味を変えずに文を圧縮したいとする。このとき、変形処理部 1 1 が使用する変形規則はすべて意味をほとんどかえずに変形するものであるとする。この場合、長さを評価の尺度とし、この長さが短くなるように変形を繰り返すと文圧縮が実現される。

【 0 0 2 8 】（ 3 ）出現頻度（または出現確率）

例えば、文章の表現を改善する推敲システムを考える。このとき、変形処理部 1 1 が使用する変形規則がすべて同義性を満足するものであるとする。この場合、推敲したいデータを、そのデータの出現（生起）確率が高くなるように変形すると非常に洗練された文章となる。

【 0 0 2 9 】もう少し簡単な例でこれを説明すると、例えば入力したデータに「データー」とあったとしよう。また、変形規則に「データー」を「データ」とする規則があったとしよう。新聞記事やコーパスなどのデータベースにより、「データー」と「データ」の出現回数を数え、「データ」の出現回数のほうが数が多い場合、「データ」のほうの評価を「データー」より高くする。

【 0 0 3 0 】また、出現頻度（または出現確率）を調べるコーパスをいろいろと変えることにより、さまざまな変換の結果を得ることができる。例えば、入力データが書き言葉のときに、コーパスとして話し言葉を用いると書き言葉の話し言葉への変形が実現される。

【 0 0 3 1 】また、入力データが法律関係の文のときに、コーパスとして平易な文章の集合を与えておくと、法律関係の難解な文章を平易な文章に変形させることが期待できる。

【 0 0 3 2 】さらにまた、ここで入力データとして適当に誰かが書いた小説の文章を入れて、コーパスとしてシ

ェークスピアの小説をいれると、シェークスピアの文体の小説が新たに完成することになる。同様に、芥川龍之介の小説を夏目漱石の文体に変形するなどといったことも可能になる。

【0033】上記の出現（生起）確率に基づく尺度は、文の正当性のチェックに使うこともできる。ただし、出現（生起）確率が尺度として強すぎる場合には、以下のような尺度を用いるとよい。

① 対象としている表現が、コーパスで1回以上出現しているか否か（これは誤植検出などにも使われる尺度である）。

② コーパスでの出現（生起）確率がある程度以上か否か。

③ コーパスでの出現（生起）確率が、環境なしでの出現確率よりも大きいかなにか。

【0034】ここで示したものは、尺度というよりは条件のようなもので、他の尺度と組み合わせて用いるとよい。他の尺度のところでも、もしその変形において、文としての正当性が保証されない場合には、ここで示した尺度を同時に用いるとよい。

【0035】図2は、変形処理部11の処理フローチャートである。変形処理部11は、まずステップS10により、言い換え対象として指定された変換対象文10を入力する。キーボードなどからの入力、ファイルからの入力、アプリケーションプログラムからの入力など、入力方法は問わない。

【0036】ステップS11では、変形規則記憶部15から変換に必要な変形規則を読み込む。既に読み込まれている場合には、ここでの読み込みは不要である。また、変形規則記憶部15に記憶されている変形規則のうち、適用システムに応じて識別子などにより特定された変形規則だけを使用する場合には、その変形規則を選択的に読み込む。

【0037】次に、ステップS12では、入力した変換対象文10を変形規則を用いて変形する。ステップS13では、変換対象文10を変形した後の表現の候補を評価処理部12へ引き渡す。このとき、変形した後の表現の候補を一つずつ評価処理部12へ渡してもよいし、複数ある場合には複数まとめて渡してもよい。

【0038】図3は、評価処理部12の処理フローチャートである。ステップS20では、評価処理部12は、変形処理部11から変換対象文10を変形した後の表現の候補を受け取る。次に、ステップS21では、評価情報記憶部16に記憶されている評価情報の中から使用する評価情報を選択する。なお、あらかじめ評価情報記憶部16に必要な評価情報だけが選択されて格納されている場合には、この処理は不要である。この評価情報は、評価処理部12から呼び出される評価関数のようなものでも、また評価関数が使用するパラメータのようなものでもよい。

【0039】ステップS22では、変形処理部11から受け取った表現の各候補を、評価情報（評価関数）を用いて評価する。ここでは、評価値が数値として算出されることになる。ステップS23では、評価結果の最もよい変形後の表現を選択し、その変形した表現を変換結果文13として出力する。

【0040】本システムは、変形処理部11が用いる変形規則および評価処理部12が用いる評価情報を、変換の目的に応じて交換または選択可能に構成することにより、各種の言語変換処理に適用することが可能である。以下、各種の処理システムへの具体的な適用例を説明する。

【0041】(A)質問応答システムへの適用例

図4は、質問応答システムへの適用例を示している。図4に示す質問応答システムでは、与えられた質問文の答えが書いてありそうな文を、新聞記事、各種論文、百科事典その他の所定の知識データベースから探し出し、その答えが書いてありそうな文と質問文との類似度が大きくなるように双方を書き換えて照合し、答えが書いてありそうな文での、質問文の疑問詞に対応している箇所を答えとして出力するといったことを行う。

【0042】この質問応答システムでは、類似度を尺度として言い換えを行っていることになる。類似度が高くなるように言い換えを行うことで質問文と回答を含むデータとの照合がしやすくなる。

【0043】本システムに入力される変換対象文10は、質問文と、回答が含まれる文の候補となる知識データベースの文である。ユーザからの質問文が、例えば「日本国の首都はどこであるか」であり、知識データベース中にある文が、「東京は日本の首都である」であったとする。

【0044】変形処理部11は、これらの二つの変換対象文10を、それぞれ質問応答システム用に用意または選択された変形規則を用いて、異なる表現に言い換える。ここで、変形規則として、図4に示すように、

①「XはYである」「YはXである」

②「日本国」「日本」

・・・

があったとする。

【0045】①の変形規則により「東京は日本の首都である」から「日本の首都は東京である」という文が生成される。また、②の変形規則により「日本国の首都はどこであるか」から「日本の首都はどこであるか」が生成される。ここでは、簡単な変形規則を例示したが、通常の変形処理では、変形された文のさらなる変形というように、多段に変形が繰り返されることになる。これらの変形した文の結果が評価処理部12に引き渡される。

【0046】評価処理部12では、質問応答システム用の評価情報（評価関数）を用いて、変形処理部11が変形した文を評価する。ここで評価の尺度が、入力した二

つの文の類似度であり、類似度が大きくなる変換が評価が高いとすると、多くの変形の中から二つの文の類似度が高いものが選ばれることになる。

【 0 0 4 7 】変形された質問文と知識データベースの文の中で類似度が最も高いものが、「日本の首都はどこであるか」と「日本の首都は東京である」であったとすると、この変換はよしとされ、これらの二つの文が変換結果文 1 3 として出力される。これらの文から「どこ」と「東京」が対応することがわかり、質問応答システムから質問文に対する回答として、「東京」または「日本国の首都は東京である」がユーザに出力されることになる。

【 0 0 4 8 】(B) 文内圧縮システムへの適用例

図 5 は、文内圧縮システムへの適用例を示している。図 5 に示す文内圧縮システムでは、要約文の作成などのために、与えられた文をできるだけ元の文の意味を保存した形で、冗長な文を短く圧縮する処理を行う。例えば、新聞記事の要約を考えた場合、評価の尺度としては、入力されたデータがより短くなるような変形をよしとする

尺度が考えられる。さらに条件として、新聞記事での出現が 1 個以上というような条件を付加してもよい。以下、具体例に従って説明する。

【 0 0 4 9 】例えば図 5 の例のように、変換対象文 1 0 として、「次の参議院選挙で A 氏を擁立することを決めた」という文が入力されたとする。変換処理部 1 1 は、この変換対象文 1 0 を、文内圧縮システム用に用意または選択された変形規則を用いて、異なる表現に言い換える。ここで、変形規則として、

「 X で Y を擁立すること」 「 X での Y の擁立」

という規則があったとすると、変換処理部 1 1 は、変換対象文 1 0 に変形規則を適用することにより、「次の参議院選挙で A 氏を擁立することを決めた」という文から「次の参議院選挙での A 氏の擁立を決めた」という文を生成する。この他にも、種々の変形規則が存在し、多くの変形された文が候補として生成されることになる。これらの文を評価処理部 1 2 に渡す。

【 0 0 5 0 】評価処理部 1 2 は、文内圧縮システム用の評価情報（評価関数）を用いて、変換処理部 1 1 が変形した文を評価する。ここで評価の尺度が、入力した文の長短であり、文の長さが短いほど評価が高いとすると、多くの変形の中から文が最も短い文が選ばれることになる。この例では、「次の参議院選挙での A 氏の擁立を決めた」の評価が高く、変換結果文 1 3 としてこの文が出力されている。

【 0 0 5 1 】(C) 推敲システムへの適用例

図 6 は、推敲システムへの適用例を示している。図 6 に示す推敲システムでは、入力した文を推敲して、より良いと考えられる表現の文に改善する処理を行う。

【 0 0 5 2 】例えば図 6 の例のように、変換対象文 1 0

として、「世界の平和・安定に貢献する」という文が入力されたとする。変換処理部 1 1 は、この変換対象文 1 0 を、推敲システム用に用意または選択された変形規則を用いて、異なる表現に言い換える。ここで、変形規則として、

「・」 「と」

という規則があったとすると、変換処理部 1 1 は、変換対象文 1 0 に変形規則を適用することにより、「世界の平和・安定に貢献する」という文から「世界の平和と安定に貢献する」という文を生成する。この他にも、種々の変形規則が存在し、多くの変形された文が候補として生成されることになる。これらの文を評価処理部 1 2 に渡す。なお、変形されなかった変換対象文 1 0 についても候補の一つとして評価処理部 1 2 に渡す。

【 0 0 5 3 】評価処理部 1 2 は、推敲システム用の評価情報（評価関数）を用いて、変換処理部 1 1 が変形した文を評価する。ここで評価の尺度が、大量の言語データ（用例、つまり実際に人々によって用いられたことのある言語表現の集合）での出現頻度もしくは出現確率が大きくなる変換をよしとするものである場合に、評価処理部 1 2 は、大量の言語データにおける「世界の平和・安定に貢献する」と「世界の平和と安定に貢献する」の生起確率を求めると、簡便な手法としては、変形した部分を含む小さい領域範囲の文字列が言語データで何回出現したかを数える。例えば「平和・安定」が 1 3 4 回、「平和と安定」が 2 3 8 2 3 回現れたとすると、「平和と安定」のほうが出現頻度が大きくこの表現のほうがより自然な表現であるとわかる。これによりこの変形はよしとされ、変換結果文 1 3 として「世界の平和と安定に貢献する」が出力される。なお、出現頻度ではなく、出現（生起）確率を計算してもよく、出現確率にしたほうが評価関数としては精度のよいものとなる。

【 0 0 5 4 】(D) 難解文変換システムへの適用例

図 7 は、難解文変換システムへの適用例を示している。図 7 に示す難解文変換システムでは、法律文章を平易な文に書き換えたり、難しい新聞の記事を小学生向けの易しい文に書き換えたりする処理を行う。

【 0 0 5 5 】例えば図 7 の例のように、変換対象文 1 0 として、「大臣を罷免する」という文が入力されたとする。変換処理部 1 1 は、この変換対象文 1 0 を、難解文変換システム用に用意または選択された変形規則を用いて、異なる表現に言い換える。ここで、変形規則として、

「罷免する」 「やめさせる」

という規則があったとすると、変換処理部 1 1 は、変換対象文 1 0 に変形規則を適用することにより、「大臣を罷免する」という文から「大臣をやめさせる」という文を生成する。この他にも、種々の変形規則が存在し、多

くの変形された文が候補として生成されることになる。これらの文を評価処理部 1 2 に渡す。なお、変形されなかった変換対象文 1 0 についても候補の一つとして評価処理部 1 2 に渡す。

【 0 0 5 6 】評価処理部 1 2 は、難解文変換システム用の評価情報（評価関数）を用いて、変形処理部 1 1 が変形した文を評価する。ここで評価の尺度が、例えば小学生向けというような低年齢層向けの文章集合での出現頻度または出現確率が大きくなる変換をよしとするものである場合に、評価処理部 1 2 は、あらかじめ定められた範囲での低年齢層向けの文章集合における「大臣を罷免する」と「大臣をやめさせる」の出現頻度を求める。簡便な手法としては、変形した部分を含む小さい領域範囲の文字列が言語データで何回出現したかを数える。「大臣をやめさせる」のほうが出現頻度が大きい場合、この表現のほうが低年齢層向けの易しい表現であるとわかる。これによりこの変形はよしとされ、変換結果文 1 3 として「大臣をやめさせる」が出力される。なお、出現頻度ではなく、出現（生起）確率を計算してもよいことは、前述した例と同様である。

【 0 0 5 7 】また、評価の尺度としては、所定の文章集合での出現頻度や出現確率に限らず、他の何らかの尺度を用いることもできる。例えば、あらかじめ単語の結び付きや、構文解析結果から得られる文法上の言い回しに対して、評価ポイントを定めておき、それを用いて評価するようなことも可能である。また、変換後の文章集合から平易さを示す何らかの尺度を学習し、それに従って評価する方法も考えられる。

【 0 0 5 8 】（ E ）個人文体変換システム

図 8 は、個人文体変換システムへの適用例を示している。図 8 に示す個人文体変換システムでは、例えば芥川龍之介の小説を、夏目漱石の文体の小説に書き換えたり、ある無名の作家の小説をシェークスピアの文体の小説に書き換えたりする処理を行う。

【 0 0 5 9 】例えば図 8 の（ 1 ）の例のように、変換対象文 1 0 として、「大臣を罷免するなどを行った」という文が入力されたとする。変形処理部 1 1 は、この変換対象文 1 0 を、個人文体変換システム用に用意または選択された変形規則を用いて、異なる表現に言い換える。ここで、「といった」という表現を多用する人の文体への変形規則として、

「するなど」 「するといったこと」
.....

という規則があったとすると、変形処理部 1 1 は、変換対象文 1 0 に変形規則を適用することにより、「大臣を罷免するなどを行った」という文から「大臣を罷免するといったことを行った」という文を生成する。この他にも、種々の変形規則が存在し、多くの変形された文が候補として生成されることになる。これらの文を評価処理部 1 2 に渡す。なお、変形されなかった変換対象文 1 0

についても候補の一つとして評価処理部 1 2 に渡す。

【 0 0 6 0 】評価処理部 1 2 は、個人文体変換システム用の評価情報（評価関数）を用いて、変形処理部 1 1 が変形した文を評価する。ここで評価の尺度が、変換目的である特定個人の文章集合での出現頻度または出現確率が高くなるような表現をよしとするものである場合に、評価処理部 1 2 は、その特定個人の文章集合における「大臣を罷免するなどを行った」という文や、「大臣を罷免するといったことを行った」という文の出現頻度を求める。なお、出現頻度は、必ずしも文全体の出現回数でなくてもよく、変形した部分を含む小さい領域範囲の文字列が文章集合の中で何回出現したかでもよい。「大臣を罷免するといったことを行った」という文の出現頻度が大きい場合、評価処理部 1 2 は、変換結果文 1 3 として「大臣を罷免するといったことを行った」を出力する。

【 0 0 6 1 】また、「であろう」を多用する人への文体への変更の場合、例えば変換規則として、

「と思われる」 「であろう」
.....

という規則を用い、評価の尺度として、その「であろう」を多用する特定個人の文章集合での出現頻度または出現確率が高くなるような表現をよしとするものを用いる。

【 0 0 6 2 】変形処理部 1 1 は、図 8 の（ 2 ）のように「大臣を罷免すると思われる」という変換対象文 1 0 を入力すると、この入力に対して変換規則を適用することにより、この文を「大臣を罷免するであろう」という表現に変形する。評価処理部 1 2 による評価によって、「大臣を罷免するであろう」という表現の評価値が最も高いことがわかると、評価処理部 1 2 はこの文を変換結果文 1 3 として出力する。

【 0 0 6 3 】なお、評価の尺度としては、所定の文章集合での出現頻度や出現確率に限らず、他の何らかの尺度を用いることができることは、前述した他のシステムの例と同様である。

【 0 0 6 4 】以上の適用例の他に、例えば入力データが書き言葉のときに、出現頻度や出現確率などによる変換候補の評価に用いるコーパスとして話し言葉の言語データを用いると、書き言葉を話し言葉へ変換するシステムが実現され、またこの逆に、変換候補の評価に用いるコーパスとして書き言葉の言語データを用いると、話し言葉を書き言葉に変換するシステムを実現することもできる。

【 0 0 6 5 】以上、種々の言語変換処理を行うシステムへの適用例を説明したが、本システムは、基本的に変形処理部 1 1 と評価処理部 1 2 の二つのモジュールによって実現することが可能である。説明を簡単にするために、主に個々の適用システムごとに、変形規則記憶部 1 5 と評価情報記憶部 1 6 とが用意される例を説明した

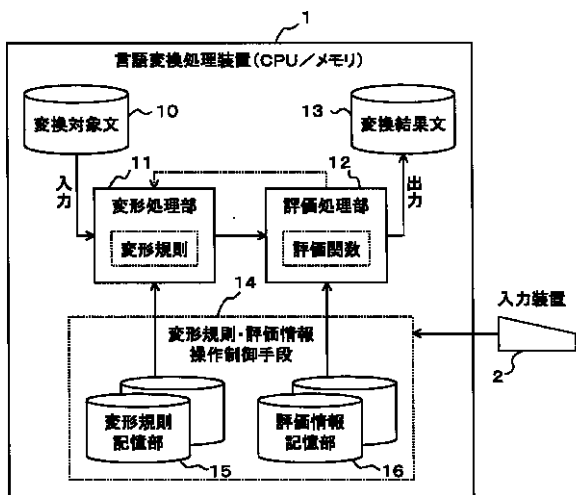
が、変形規則記憶部 1 5 の変形規則または評価情報記憶部 1 6 の評価関数などの一部または全部を、複数の適用システム間で共用することも可能である。特に変形規則は、ほとんどのケースが、ある文または文節などを同義の他の文または文節に変換するものであるから、複数の適用システム間で共用することができる。もちろん、変形規則を複数の適用システム間で共用しないで、それぞれの適用システムに特化したものとしてもよい。それぞれの適用システムに適した変形規則を用いることにより、文の変形処理および変形後の文の評価処理における処理コストを削減することができる。

【 0 0 6 6 】

【発明の効果】以上説明したように、本発明によれば、種々の言語変換処理を行うシステムを構築する場合に、変形規則および評価情報（評価関数）を適用システムに応じて用意するだけで、新システムの構築を実現することが可能になる。また、言語変換のインターフェースの統一化が可能であり、利用しやすいユーザ・インターフェースおよびアプリケーション・インターフェースの提供が可能になる。

【図面の簡単な説明】

【図 1】



【図 1】本発明のシステム構成例を示す図である。

【図 2】変形処理部の処理フローチャートである。

【図 3】評価処理部の処理フローチャートである。

【図 4】質問応答システムへの適用例を示す図である。

【図 5】文内圧縮システムへの適用例を示す図である。

【図 6】推敲システムへの適用例を示す図である。

【図 7】難解文変換システムへの適用例を示す図である。

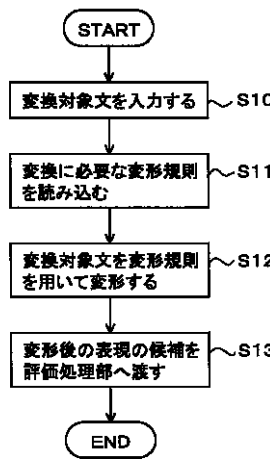
【図 8】個人文体変換システムへの適用例を示す図である。

【符号の説明】

- 1 言語変換処理装置
- 2 入力装置
- 10 変換対象文
- 11 変形処理部
- 12 評価処理部
- 13 変換結果文
- 14 変形規則・評価情報操作制御手段
- 15 変形規則記憶部
- 20 16 評価情報記憶部

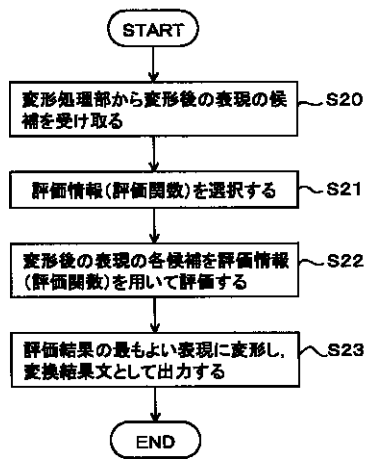
【図 2】

変形処理部の処理フローチャート

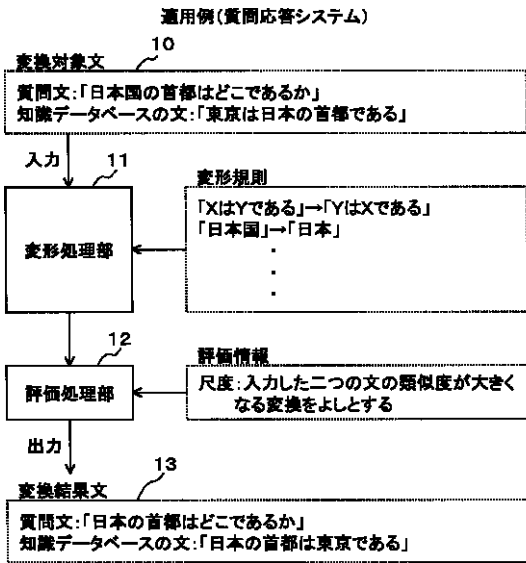


【図 3】

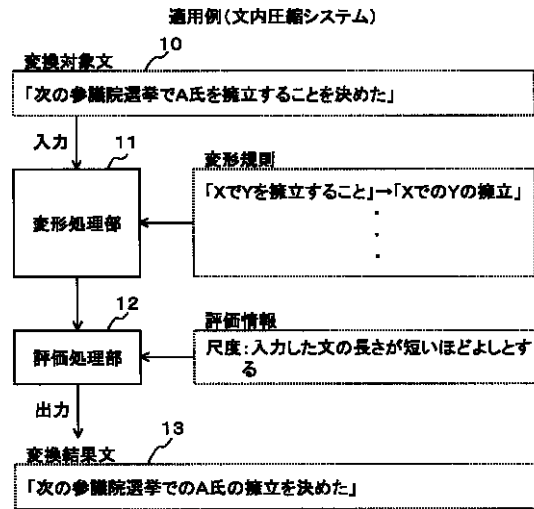
評価処理部の処理フローチャート



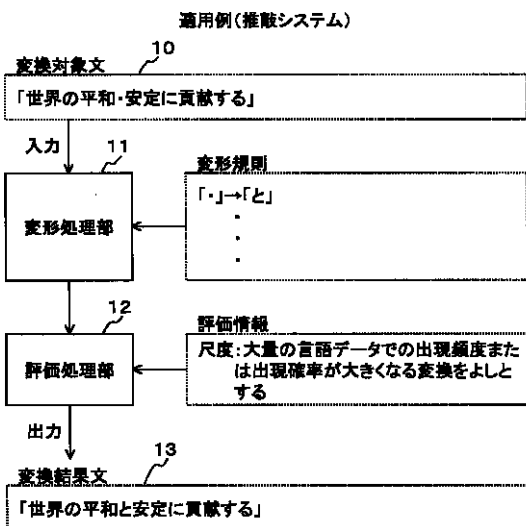
【 図 4 】



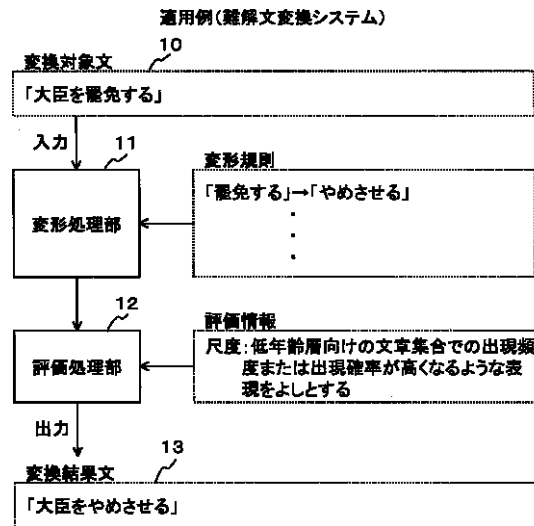
【 図 5 】



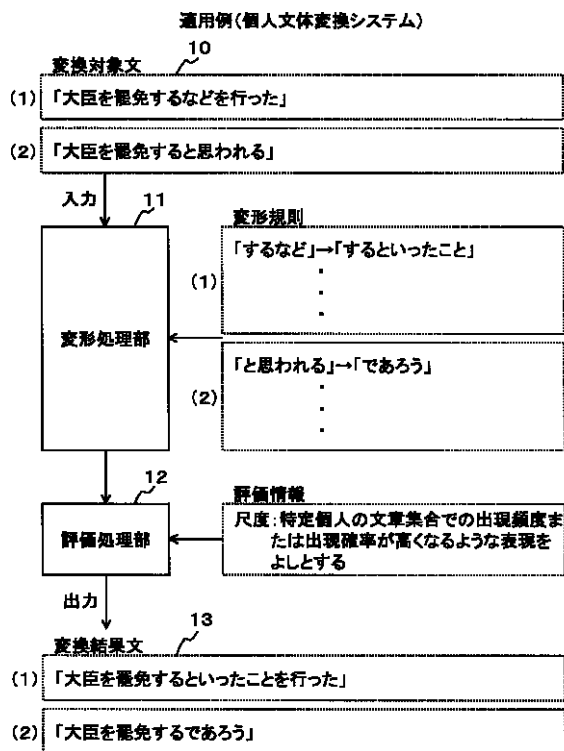
【 図 6 】



【 図 7 】



【 図 8 】



【 手続補正書 】

【 提出日 】 平成 1 4 年 7 月 1 5 日 (2 0 0 2 . 7 . 1 5)

【 手続補正 1 】

【 補正対象書類名 】 明細書

【 補正対象項目名 】 特許請求の範囲

【 補正方法 】 変更

【 補正内容 】

【 特許請求の範囲 】

【 請求項 1 】 ある自然言語で記述された文字列を、同一の自然言語で記述された他の表現による文字列に変換するシステムにおいて、前記自然言語の文字列に関する変形の規則を記憶する変形規則記憶手段と、文字列の変換目的に応じて用意される評価情報であって、文字列を変形した結果が目的とするふさわしい変換であるかどうかを評価するための尺度を与える評価関数または評価規則の評価情報を記憶する評価情報記憶手段と、自然言語で記述された変換対象の文字列を入力する入力手段と、前記入力された文字列を前記変形規則記憶手段に記憶された変形の規則を用いて変形し、変換の候補を生成する、変換目的の異なる複数の変換方式に対応可能な変換処理手段と、前記変形処理手段によって生成された変換の候補を、前記評価情報記憶手段に記憶された評価関数または評価規則を用いて評価し、評価結果のよい表現を

選択する、変換目的の異なる複数の変換方式に対応可能な評価処理手段と、前記変形処理手段による変形および前記評価処理手段による評価結果から、評価の高い文字列の変換結果を出力する出力手段とを備え、かつ、前記変形処理手段が用いる変形規則および前記評価処理手段が用いる評価関数または評価規則は、変換の目的に応じて交換または選択可能に構成されていることを特徴とする言語変換処理統一システム。

【 請求項 2 】 前記変換対象となる文字列は、質問応答システムにおける質問文と質問文に対する回答を含む文の候補の文字列であり、前記変形規則は、第 1 の文字列を同義の第 2 の文字列に言い換えるパターン規則であり、前記評価関数または評価規則による評価の尺度は、質問文と回答を含む文の候補との類似度の大小を含み、類似度の大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【 請求項 3 】 前記文字列の変換は、入力した文字列から冗長な表現を修正して短くした文字列を生成する文内圧縮システムのための変換であり、前記変形規則は、第 1 の文字列を第 2 の文字列に言い換えるパターン規則であり、前記評価関数または評価規則による評価の尺度は、変形後の文字列の長短を含み、変形後の文字列が短

いものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【請求項 4】 前記文字列の変換は、入力した文字列の表現を推敲する推敲システムのための変換であり、前記変換規則は、第 1 の文字列を同義の第 2 の文字列に言い換えるパターンの規則であり、前記評価関数または評価規則による評価の尺度は、大量の用例に関する言語データ中に現れる出現頻度または出現確率を含み、出現頻度または出現確率が大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【請求項 5】 前記文字列の変換は、入力した難解文の文字列を平易文に変換する難解文変換システムのための変換であり、前記変換規則は、第 1 の文字列を同義の第 2 の文字列に言い換えるパターンの規則であり、前記評

価関数または評価規則による評価の尺度は、平易な文章集合からなる大量の用例における出現頻度または出現確率を含み、出現頻度または出現確率が大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。

【請求項 6】 前記文字列の変換は、入力した文字列の表現をある特定の個人の文体に変換する個人文体変換システムのための変換であり、前記変換規則は、第 1 の文字列を同義の第 2 の文字列に言い換えるパターンの規則であり、前記評価関数または評価規則による評価の尺度は、前記特定の個人の文章集合からなる大量の用例における出現頻度または出現確率を含み、出現頻度または出現確率が大きいものに高い評価を与えるものであることを特徴とする請求項 1 記載の言語変換処理統一システム。