(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号

特開2003－76687

（Ｐ２００３－７６６８７Ａ）

(43)公開日　平成15年３月14日(2003.3.14)

(51)Int.Cl.[7]	識別記号	ＦＩ	ﾃｰﾏｺｰﾄﾞ (参考)
G06F 17/28		G06F 17/28	P 5B091

審査請求　有　請求項の数４　ＯＬ　（全12頁）

(21)出願番号　特願2001－268513(Ｐ2001－268513)

(22)出願日　平成13年９月５日(2001.9.5)

特許法第30条第１項適用申請有り　平成13年３月30日　言語処理学会開催の「言語処理学会第７回年次大会」において文書をもって発表

(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ー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）

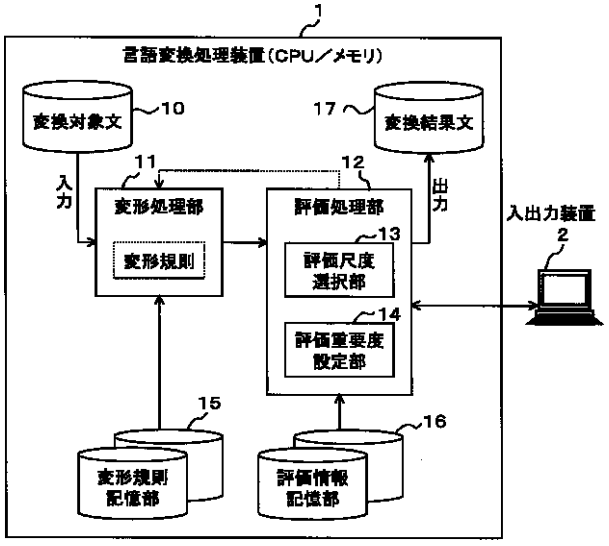
Ｆターム(参考)　5B091 AA15 CA21 CC03 CC15 DA06

(54)【発明の名称】複数尺度の利用による言語変換処理システムおよびその処理プログラム

(57)【要約】

【課題】　複数種類の言い換えが必要な文または文章を，目的とする文または文章に簡単に変換することができるシステムを提供する。

【解決手段】　変形処理部１１は，変換対象文１０を入力すると，変形規則記憶部１５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１２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複数の評価尺度を用いて評価し，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１７として出力する。評価尺度は，評価尺度選択部１３によって選択することができ，また選択した評価尺度の重要度は，評価重要度設定部１４によって設定することができる。



1

【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他の表現による文字列に変換するシステムであって，自然言語で記述された変換対象の文字列を入力する入力手段と，前記入力された文字列を所定の変形規則または所定の変形処理プログラムを用いて変形し，変換の候補を生成する変形処理手段と，文字列を変形した結果の表現が目的とする表現になっているかどうかを評価するためのＮ種類（Ｎ≧２）の評価尺度を組み合わせ，各評価尺度に応じて前記変形処理手段により生成された変換の候補を評価し，Ｎ種類の評価尺度による総合的な評価結果のよい表現を選択する評価処理手段と，前記選択された表現の変換結果を，目的とする表現に変換された文字列として出力する出力手段とを備えることを特徴とする複数尺度の利用による言語変換処理システム。

【請求項２】　前記評価処理手段が変換の候補の評価に用いるＮ種類の評価尺度を，あらかじめ用意されたＭ種類（Ｍ≧Ｎ）の評価尺度の中から外部からの指定により選択する手段を備えることを特徴とする請求項１記載の複数尺度の利用による言語変換処理システム。

【請求項３】　前記評価処理手段が変換の候補の評価に用いるＮ種類の評価尺度に対して，各評価尺度の重要度に関する指定情報を入力する手段を備え，前記評価処理手段は，入力された指定情報に基づいて，個々の評価尺度に対する重要度に応じた評価結果から総合的な評価を行うことを特徴とする請求項１または請求項２記載の複数尺度の利用による言語変換処理システム。

【請求項４】　前記Ｎ種類の評価尺度の少なくとも一つは，変形後の文字列の長短，大量の用例に関する言語データ中に現れる出現頻度もしくは出現確率，所定の平易な文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，口語もしくは文章語で表現された大量の文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，特定の個人の文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，または，変換対象となっている複数の文字列の類似度のいずれかであることを特徴とする請求項１，請求項２または請求項３記載の複数尺度の利用による言語変換処理システム。

【請求項５】　コンピュータによって，ある自然言語で記述された文字列を他の表現による文字列に変換するためのプログラムであって，自然言語で記述された変換対象の文字列を入力する処理と，前記入力された文字列を所定の変形規則または所定の処理手続きによって変形し，変換の候補を生成する処理と，文字列を変形した結果の表現が目的とする表現になっているかどうかを評価するためのＮ種類（Ｎ≧２）の評価尺度を組み合わせ，各評価尺度に応じて前記変形によって生成された変換の候補を評価し，Ｎ種類の評価尺度による総合的な評価結果のよい表現を選択する処理と，前記選択された表現の変換結果を，目的とする表現に変換された文字列として出力する処理とを，コンピュータに実行させるための複数尺度の利用による言語変換処理プログラム。

2

【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，同一または他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扱うことができるようにした複数尺度の利用による言語変換処理システムおよびその処理プログラ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内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，各システムごとにそれぞれ個別に独自の変換の処理ロジックを用いているため，例えば口語で表現された文章の要約を作成するというような場合には，まず口語文章語変換システムにより口語文を通常の文章語による文に変換し，その結果について要約文生成システムにより要約を生成するというような処理が必要であった。

【０００６】また，与えられた文章を推敲し，それについて要約を作成するというような場合にも，まず推敲システムにより文章を推敲し，その後に要約文生成システムによって要約を生成するか，最初に要約文生成システムによって要約を生成し，その結果を推敲システムを用いて推敲するという処理が必要であった。このとき，文章の変換は一律に行われ，要約のほうを推敲よりも重視するとか，これとは反対に推敲のほうを要約よりも重視するということはできなかった。

【０００７】本発明は上記問題点の解決を図り，複数種類の言い換えが必要な文または文章を，目的とする文または文章に簡単に変換することができるシステムを提供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，同一または異なる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，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，入力された文字列を所定の変形規則または所定の変形処理プログラムを用いて変形し，変換の候補を生成する変形処理手段と，文字列を変形した結果の表現が目的とする表現になっているかどうかを評価するためのN種類（N≧２）の評価尺度を組み合わせ，各評価尺度に応じて変形処理手段により生成された変換の候補を評価し，N種類の評価尺度による総合的な評価結果のよい表現を選択する評価処理手段と，選択された表現の変換結果を，目的とする表現に変換された文字列として出力する出力手段とを備え，複数の評価尺度を同時に利用して言語変換処理を行う。

【０００９】また，評価処理手段が変換の候補の評価に用いるN種類の評価尺度は，あらかじめ用意されたM種類（M≧N）の評価尺度の中から，ユーザまたはアプリケーションプログラム等からの指定により選択する手段を持つ。

【００１０】さらに，評価処理手段が変換の候補の評価に用いるN種類の評価尺度に対して，各評価尺度の重要度に関する指定情報を入力する手段を設け，評価処理手段は，入力された指定情報に基づいて，個々の評価尺度に対する重要度に応じた評価結果から総合的な評価を行うようにすることもできる。

【００１１】N種類の評価尺度としては，変形後の文字列の長短，大量の用例に関する言語データ（コーパス）中に現れる出現頻度もしくは出現確率，所定の平易な文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，口語もしくは文章語で表現された大量の文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，特定の個人の文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，または，変換対象となっている複数の文字列の類似度などを用いることができる。

【００１２】評価尺度として，変形後の文字列の長短を用い，変形後の文字列が短いものに高い評価を与えれば，冗長な表現を短くした変換文字列を生成することができる。また，評価尺度として，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用い，その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れば，一般によく使われる文または文章になるように推敲した変換文字列を得ることができる。

【００１３】評価尺度として，所定の平易な文章集合からなる言語データ中に現れる出現頻度または出現確率を用い，その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れば，法律文などの難解な文を平易な文に変換した文字列を得ることができる。

【００１４】また，評価尺度として，口語で表現された大量の文章集合からなる言語データ中に現れる出現頻度または出現確率を用い，その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れば，文章語を口語表現に変換した文字列を生成することができる。これとは逆に，評価尺度として，文章語で表現された大量の文章集合からなる言語データ中に現れる出現頻度または出現確率を用い，その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れば，口語表現を文章語の表現に変換した文字列を生成することができる。

【００１５】さらに，評価尺度として，例えば夏目漱石とか芥川龍之介といった特定の個人の文章集合からなる言語データ中に現れる出現頻度または出現確率を用い，その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れば，与えられた文章を夏目漱石の文体もしくは芥川龍之介の文体といった特定の個人の文体に変換することができる。

【００１６】また，評価尺度として，変換対象となっている複数の文字列の類似度を用いれば，与えられた複数の文を比較する場合に，単なる表現形式ではなく，実質的な内容も考慮に入れた類似度を比較できるような文に変換することができるようになる。

【００１７】本発明では，特に以上のような評価尺度を複数組み合わせて用いることができる。したがって，与えられた文章を推敲し，それについて要約を作成するというような文字列の変換や，難解な文を易しい文に変換し，しかもそれを特定の個人の文体で表現するというような変換を，一度で行うことができるようになる。また，各評価尺度の重要度を随時変えることにより，目的とする変換文字列が得られるように調整することができる。

【００１８】以上の手段は，コンピュータと，そのコンピュータにインストールされ実行されるソフトウェアプログラムとによって実現することができ，そのプログラムは，コンピュータが読み取り可能な可搬媒体メモリ，半導体メモリ，ハードディスク等の適当な記録媒体に格納することができる。

【００１９】

【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はディスプレイ，キーボードその他の入出力装置を表す。

【００２０】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１７は，本システムの出力であって，変換対象文１０を同一の種類または異なる種類の自然言語で言い換えたものである。

【００２１】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１１と評価処理部１２とから構成される。変形処理部１１は，変形規則記憶部１５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１２は，変換の候補のよさを，あらかじめ評価情報記憶部１６に記憶されている複数の評価尺度（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。

【００２２】評価処理部１２は，変換候補の評価に用いる評価尺度を入出力装置２からの指定によって選択する評価尺度選択部１３と，選択された各評価尺度の重要度に関する指定情報を入出力装置２から入力し設定する評価重要度設定部１４とを持ち，評価尺度選択部１３により選択された評価尺度と，評価重要度設定部１４によって設定された各評価尺度の重要度とから，個々の評価尺度に対する重要度に応じた変換候補の総合的な評価を行い，総合的な評価結果のよい変換候補を選択する。

【００２３】言語変換処理装置１の動作は，以下のとおりである。変換対象文１０が入力されると，変形処理部１１は，変形規則を用いて変換の候補を挙げ，評価処理部１２は，変換の候補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，その結果を変換結果文１７として出力する。

【００２４】変形規則記憶部１５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。例えば，同義性を満足する変形規則を自動獲得する方法の例としては，次のような方法を挙げることができる。異なる複数の辞書の同じ項目の定義文を照合し，その照合結果から変形規則を得る。例えば「あべこべ」という語の定義文を考えてみる。大辞林（三省堂）では，「あべこべ」の説明文が，「順序・位置などの関係がさかさまに入れかわっていること。」となっており，岩波国語辞典では，「順序・位置・関係がひっくり返っていること。」となっている。これを適当に照合すると，「関係が」と「こと。」が一致し，その間の「さかさまに入れかわっている」と「ひっくり返っている」が同義表現として機械的に獲得される。

【００２５】変換の候補を評価する評価尺度（評価関数）の評価情報は，扱う問題ごとに適正なものが複数種類，あらかじめ評価情報記憶部１６に用意される。評価尺度としての評価情報は，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１２で用いる評価尺度の例としては，以下のようなものが考えられる。

【００２６】（１）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１１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。

【００２７】（２）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１１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。

【００２８】（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を考える。このとき，変形処理部１１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。

【００２９】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。

【００３０】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。

【００３１】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。

【００３２】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。さらに，評価尺度として，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。

【００３４】評価尺度を条件のようにして用いることもできる。条件のような尺度として，例えば「２１世紀」というような特定の語を使うことに高い評価を与えたり，起承転結を満足する文章構成をとる変換に高い評価を与えたり，係り先未決定文節数が８程度以上である変換は評価を低くするということが考えられる。また，英語文でＲやＬを含む発音しにくい単語をあまり使わないというような尺度も考えられる。

【００３５】以上の評価尺度を複数組み合わせて用いることにより，多種多様な文字列の変換を実現することができる。本発明は，複数の評価尺度を組み合わせて望ましい変換結果を得ることができるようにしたものであるが，以下では，本発明の理解を容易にするために，各種の評価尺度を単独で用いた場合の具体例について説明する。

【００３６】（Ａ）文内圧縮の変換例

図２は，文内圧縮（要約文生成）の変換例を示している。図２に示す変換では，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。

【００３７】例えば図２の例のように，変換対象文１０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。変形処理部１１は，この変換対象文１０を，変形規則記憶部１５にあらかじめ用意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，変形処理部１１は，変換対象文１０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１２に渡す。

【００３８】評価処理部１２は，文内圧縮用の評価尺度（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１７としてこの文が出力されている。

【００３９】（Ｂ）文章推敲の変換例

図３は，文章推敲のための変換例を示している。図３に示す文章推敲では，入力した文または文章を推敲して，より良いと考えられる表現の文または文章に改善する処理を行う。

【００４０】例えば図３の例のように，変換対象文１０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。変形処理部１１は，この変換対象文１０を，変形規則記憶部１５にあらかじめ用意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，変形処理部１１は，変換対象文１０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１２に渡す。なお，変形されなかった変換対象文１０についても候補の一つとして評価処理部１２に渡す。

【００４１】評価処理部１２は，文章推敲用の評価尺度（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，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１２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。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１３４回，「平和と安定」が２３８２３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１７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。

【００４２】（Ｃ）難解文の平易文への変換例

図４は，難解文を平易文に変換した変換例を示している。図４に示す難解文の平易文への変換では，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。

【００４３】例えば図４の例のように，変換対象文１０として，「大臣を罷免する」という文が入力されたとする。変形処理部１１は，この変換対象文１０を，変形規則記憶部１５にあらかじめ用意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，変形処理部１１は，変換対象文１０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１２に渡す。なお，変形されなかった変換対象文１０についても候補の一つとして評価処理部１２に渡す。

【００４４】評価処理部１２は，難解文変換用の評価尺度（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，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１２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１７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。

【００４５】（Ｄ）特定個人文体への変換例

図５は，特定の個人文体への変換例を示している。図５に示す特定個人文体への変換では，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。

【００４６】例えば図５の（１）の例のように，変換対象文１０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１１は，この変換対象文１０を，変形規則記憶部１５にあらかじめ用意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，

「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１１は，変換対象文１０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１２に渡す。なお，変形されなかった変換対象文１０についても候補の一つとして評価処理部１２に渡す。

【００４７】評価処理部１２は，特定個人文体への変換用の評価尺度（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，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１２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１２は，変換結果文１７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。

【００４８】また，例えば変形規則として，

「と思われる」→「であろう」

・・・・

という規則をあったとする。ある文章を，「であろう」を多用する人の文体に変換することを考える。この場合，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。

【００４９】変形処理部１１は，図５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１０を入力すると，この入力に対して変形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１２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１２はこの文を変換結果文１７として出力する。

【００５０】（Ｅ）質問応答システムのための変換例

図６は，質問応答システムのための変換例を示している。図６に示す変換では，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。

【００５１】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。

【００５２】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

11

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。

【００５３】変形処理部１１は，これらの二つの変換対象文１０を，それぞれ変形規則記憶部１５にあらかじめ用意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６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。

【００５４】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。

【００５５】評価処理部１２では，質問応答システム用の評価尺度（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。

【００５６】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１７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。

【００５７】質問応答システムにおいて類似度を尺度として言い換えを行った例を説明したが，同様に情報検索においても類似度を尺度とした変換を利用することができる。この場合，検索のクエリと検索される記事との類似度が高くなるように言い換えてから，クエリと記事との類似度を求める。

【００５８】照応の問題でも，「近くの大きな杉の木の根元にある穴」と「杉の木の根元の穴」の同一性もしくは包含関係が判定できないと照応を解決できないというのがあるが，類似度を尺度として両者を言い換え，例えば「近くの大きな杉の木の根元の穴」と「杉の木の根元の穴」になった場合，後者が前者に含まれることが分かり，後者が前者を指示可能となる。

【００５９】以上の変換例の他に，例えば入力データが書き言葉（文章語）のときに，出現頻度や出現確率など

12

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（口語）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。

【００６０】また，ある自然言語で記述された文を他の自然言語で記述された文に変換する機械翻訳にも，次のように適用することができる。機械翻訳への適用の場合，変形規則記憶部１５に記憶する変形規則として翻訳規則を入れ，評価処理部１２では，ターゲット側の自然言語の言語コーパスを用いて，その言語コーパスにおける出現頻度または出現確率などを評価尺度として用いる。

【００６１】本発明では，以上説明したような個々の評価尺度を同時に複数利用して与えられた文字列を変換することを可能にする。さらに，どの評価尺度を重視した変換を行うかについても指定可能にする。

【００６２】例えばある文書を文内圧縮（要約）し，かつ推敲することを考える。文内圧縮（要約）の評価尺度は，文の長短であった。また，推敲の評価尺度は，大量な用例の言語コーパスでの出現頻度（出現確率）であった。

【００６３】このとき，ユーザは，要約のほうを推敲よりも重視したいとする。ユーザは，評価尺度として文の長短と，大量な用例の言語コーパスでの出現頻度（出現確率）を選択し，文の長短の評価尺度としての重みを大きく設定する。例えば評価尺度を，①「長さを短くする」，②「長さが同じ場合には言語コーパスでの出現確率を高くする」というように設定してもよい。

【００６４】また，ユーザが推敲のほうを要約よりも重視したいというような場合には，例えば評価尺度を，①「言語コーパスでの出現確率を高くする」②「言語コーパスでの確率が同じ場合には長さを短くする」と設定すればよい。

【００６５】さらに，はっきりとどちらを重視するという指定ではなく，適当な比率ｔを用いて，評価尺度を「（文の長さの逆数）×（コーパスでの出現確率）^{t}」と設定するというような実施も可能である。

【００６６】同様に，機械翻訳し，その要約を生成するといった変換や，口語表現を文章語表現に変換し，さらにその要約を生成するといった変換についても，それぞれに適した評価尺度の選択によって，容易に実現することができる。前者の変換の場合，機械翻訳における目的言語のコーパスでの出現確率，要約での文の長さを評価尺度として用いればよい。また，後者の変換では，文章語のコーパスでの出現確率と，文の長さを評価尺度として選択すればよい。

【００６７】「要約」と「口語文章語変換」と「推敲」を同時に行う変換の例を説明する。口語文章語変換と推

敵は，文章語のコーパスの確率を評価尺度として用いる。これらは，同じ評価尺度を利用することも可能である。要約は，文の長さを評価尺度とする。ここでは，両方の尺度を，以下のような重みづけで用いることにする。ｔはユーザが指定することができる重要度のパラメータである。

【００６８】・総合評価尺度：（文の長さの逆数）×（コーパスでの出現確率）^{t}

変換対象文１０として，講演会での次のような話し言葉が入力されたとする。

（入力）「今日はえー単語を意味でソートすることについてお話しします。」

変形規則としては，次のような規則が変形規則記憶部１５に登録されていたとする。

【００６９】「えー」→＊（削除）
「ＡをＢでＣする」→「ＡのＢのＣ」
「お話しします」→「お話しする」
「お話しする」→「述べる」
「Ａについて述べる」→「Ａを述べる」
「ＡのＢ」→「ＡＢ」
「今日は」→「今日」
…　　…

これらの変形規則を用いることで，要約と口語文章語変換を同時に行うことができる。これらの変形規則により多くの変換の候補が作成されることになり，その変換の候補の選択は，要約と口語文章語変換の各評価尺度を用いて行われ，総合的に評価結果のよい候補が変換結果文１７として出力される。

【００７０】「今日はえー単語を意味でソートすることについてお話しします。」という変換対象文１０に対して，上記変形規則を適用することにより，多数の変換の候補が得られるが，ここで，要約の評価尺度だけを使った場合には，

「今日単語意味ソートを述べる。」

という変換の候補が，変換結果文１７として選択される。この文は「今日」「単語意味ソート」の部分が読みにくい文となっている。

【００７１】一方，口語文章語変換と推敲の評価尺度だけを使った場合には，

「今日は単語を意味でソートすることについて述べる。」

という変換の候補が，変換結果文１７として選択される。この文は，それほど短い文になっていない。

【００７２】ここで，要約と口語文章語変換・推敲の両方の評価尺度を同時に用いることで

「今日は単語の意味ソートを述べる。」

という変換の候補が，変換結果文１７として選択され，簡潔で適切な表現が得られることになる。

【００７３】図７は，図１に示す言語変換処理装置１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。

【００７４】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。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。ここで変形規則は，適用可能なものを繰り返し適用することができ，変形規則の適用によって多数の変換の候補が生成されることになる。それらの変換の候補を作業用メモリに蓄積する。

【００７５】ステップＳ１３では，ユーザに使用する評価尺度を選択させるために，評価尺度の選択画面を表示する。図８に，評価尺度選択画面の例を示す。評価尺度の選択項目として，「短い文」「よく使われる表現」「平易な文」「著名な作家の文体」「口語の表現」「文章語の表現」「二つの文の類似度」…などがあり，これらは複数の選択が可能である。「著名な作家の文体」が選択された場合には，さらに作家名の入力が要求される。

【００７６】ステップＳ１４では，選択終了のボタンがマウス等によりクリックされると，そのときに選ばれている評価尺度の選択情報を入力する。図８の例では，「短い文」「著名な作家（夏目漱石）の文体」が選択されており，ユーザは，変換対象文１０を夏目漱石の文体で短い文（要約）に変換することを指定している。

【００７７】次に，ステップＳ１５では，選択された各評価尺度の重要度指定画面を表示する。図９に，評価尺度の重要度指定画面の例を示す。図９（Ａ）の画面は，評価尺度として「短い文」「よく使われる表現」「平易な文」が選択された場合の例であり，これらの評価尺度に対応するつまみをマウス等によりスライドさせる操作によって，各評価尺度の重要度を指定することができるようになっている。

【００７８】また，図９（Ｂ）の画面は，評価尺度として「短い文」と「平易な文」の二つが選択されたときの重要度指定画面であり，つまみを左側へ動かせば，相対的に「短い文」であることが重視され，つまみを右側へ動かせば，「平易な文」であることが重視される。

【００７９】ステップＳ１６では，図９（Ａ）または（Ｂ）の画面において「ＯＫ」のボタンが押されたときに，その時のつまみの位置から各評価尺度に対する重要度を決定する。

【００８０】続いてステップＳ１７では，ステップＳ１２で生成された変形後の表現の各候補について，ステップＳ１４で入力した各評価尺度（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。

【００８１】次に，ステップＳ１８では，ステップＳ１

６で入力した各評価尺度の重要度に従って総合的な評価結果を算出する。例えば各評価尺度による評価値がｖ１，ｖ２，…，ｖｎであり，各評価尺度の重要度がｔ１，ｔ２，…，ｔｎであったときに，総合的な評価値Ｖを，

Ｖ＝ｔ１×ｖ１＋ｔ２×ｖ２＋…＋ｔｎ×ｖｎ

というような式によって計算してもよいし，

$V = v1^{t1} \times v2^{t2} \times \cdots \times vn^{tn}$

というような式によって計算してもよい。

【００８２】ステップＳ１９では，総合的な評価値Ｖの最も大きい変形後の表現（変換の候補）を選択し，それを変換結果文１７として出力する。その後，ステップＳ２０では，評価尺度を変更するかどうかをユーザに問い合わせる。ユーザは，出力された変換結果文１７が満足できる表現になっていれば，評価尺度の変更は指示しないで，処理を終了させる。もし，出力された変換結果文１７が満足できる表現になっていなければ，評価尺度の変更を指示することができる。

【００８３】ステップＳ２１では，評価尺度の変更の指示があると，ステップＳ１３へ制御を戻し，評価尺度の選択処理から変換結果文１７の出力まで同様に処理を繰り返す。これにより，ユーザは望ましい結果が得られるまで，評価尺度またはその重要度を変えて，対話的に変換の処理を進めることができる。

【００８４】

【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同義表現の変形規則と評価尺度を用いて行い，各種の文（文章）の言い換えを行うことができるようになる。このとき，言い換えの目的に応じて変える部分は，評価尺度の部分だけである。このため複数の変換内容を含む変換，例えば要約と推敲を同時に行うというような変換を，複数の評価尺度を選択し，複合した意味合いの尺度を用いることで簡単に実現することができる。これはユーザにとって好ましい変換を行うことができるという利点があり，システム設計においても柔軟で多様な言語変換機能を簡易な構成で提供することができるという大きな利点がある。また，出力の精度も，要約生成システムにより要約し終わったものを推敲システムにより推敲するという２段階の処理構成をとる場合よりも高くなることが期待できる。

【図面の簡単な説明】

【図１】本発明のシステム構成例を示す図である。

【図２】文内圧縮の変換例を示す図である。

【図３】文章推敲の変換例を示す図である。

【図４】難解文を平易文に変換した変換例を示す図である。

【図５】特定の個人文体への変換例を示す図である。

【図６】質問応答システムのための変換例を示す図である。

【図７】言語変換処理の処理フローチャートである。

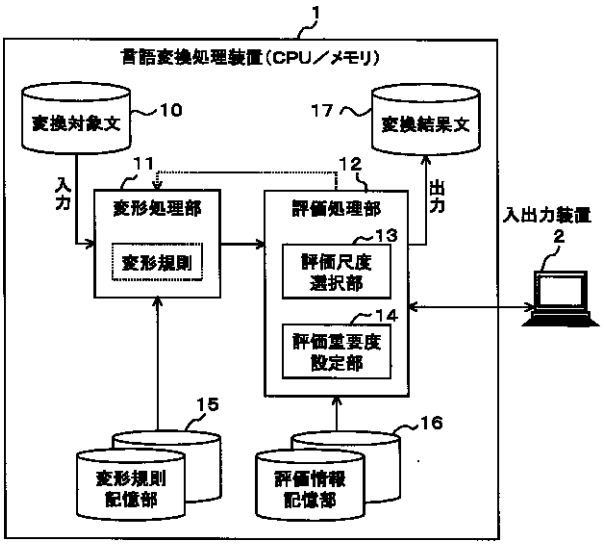
【図８】評価尺度選択画面の例を示す図である。

【図９】評価尺度の重要度指定画面の例を示す図である。

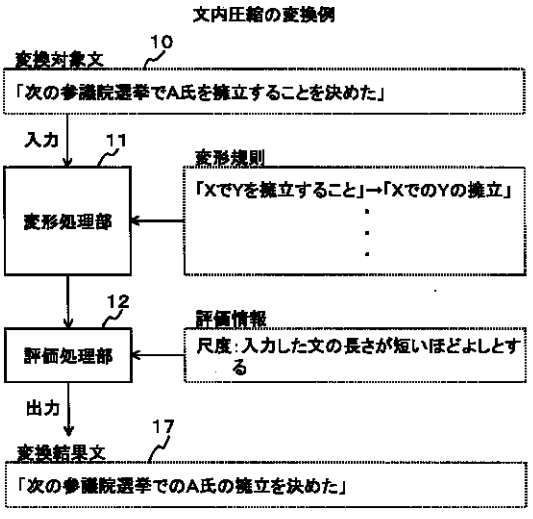
【符号の説明】

１　言語変換処理装置
２　入出力装置
１０　変換対象文
１１　変形処理部
１２　評価処理部
１３　評価尺度選択部
１４　評価重要度設定部
１５　変形規則記憶部
１６　評価情報記憶部
１７　変換結果文

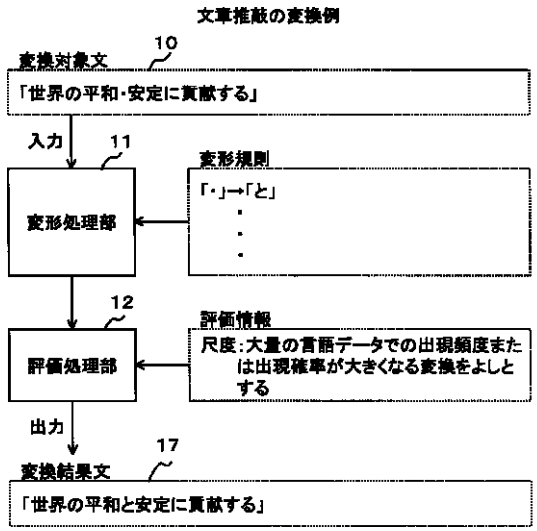
【図１】



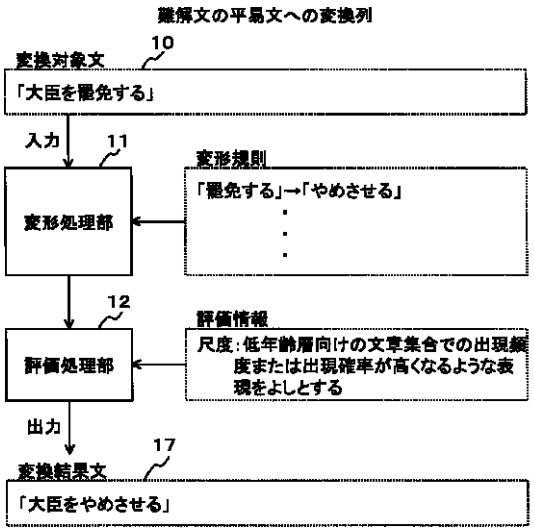
【図２】



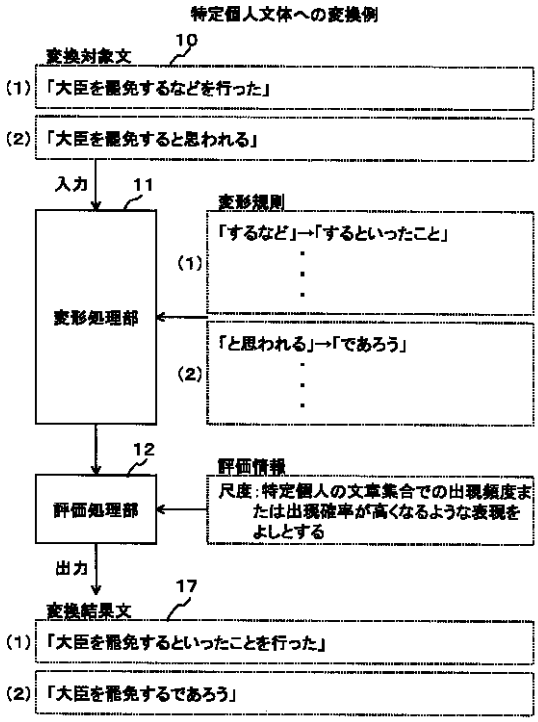
【図３】


文章推敲の変換例
変換対象文
10
「世界の平和・安定に貢献する」
入力
11
変形処理部
変形規則
「・」→「と」
12
評価処理部
評価情報
尺度：大量の言語データ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
出力
17
変換結果文
「世界の平和と安定に貢献する」

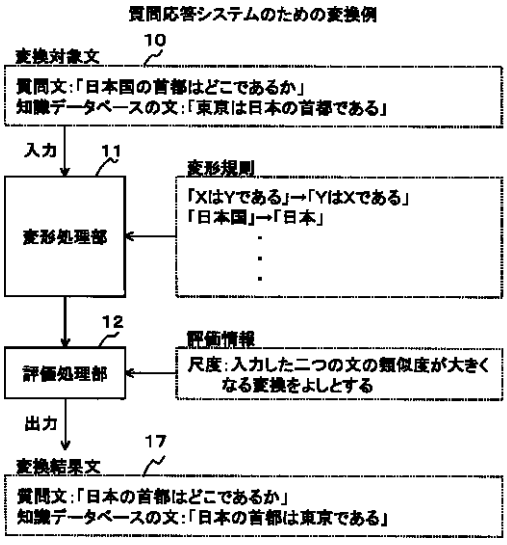
【図４】


難解文の平易文への変換例
変換対象文
10
「大臣を罷免する」
入力
11
変形処理部
変形規則
「罷免する」→「やめさせる」
12
評価処理部
評価情報
尺度：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高くなるような表現をよしとする
出力
17
変換結果文
「大臣をやめさせる」

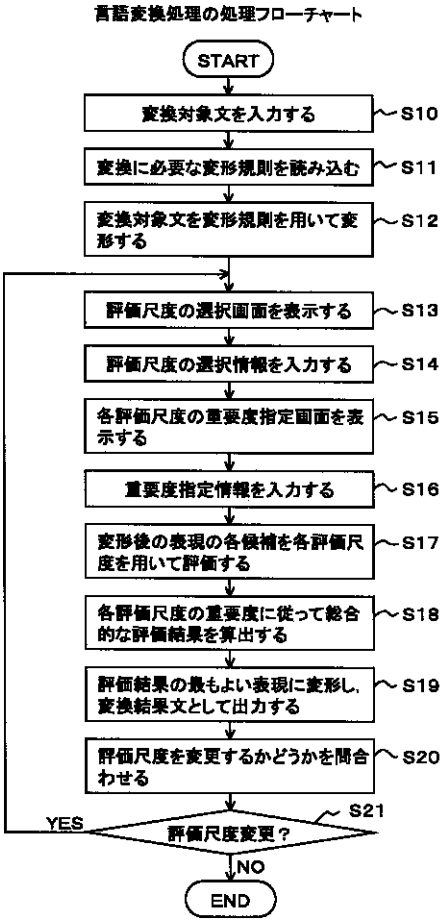
【図５】


特定個人文体への変換例
変換対象文
10
(1)「大臣を罷免するなどを行った」
(2)「大臣を罷免すると思われる」
入力
11
変形処理部
変形規則
(1)「するなど」→「するといったこと」
(2)「と思われる」→「であろう」
12
評価処理部
評価情報
尺度：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
出力
17
変換結果文
(1)「大臣を罷免するといったことを行った」
(2)「大臣を罷免するであろう」

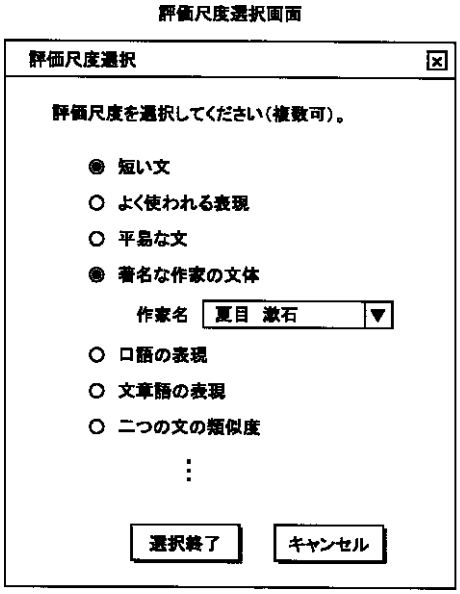
【図６】


質問応答システムのための変換例
変換対象文
10
質問文：「日本国の首都はどこであるか」
知識データベースの文：「東京は日本の首都である」
入力
11
変形処理部
変形規則
「XはYである」→「YはXである」
「日本国」→「日本」
12
評価処理部
評価情報
尺度：入力した二つの文の類似度が大きくなる変換をよしとする
出力
17
変換結果文
質問文：「日本の首都はどこであるか」
知識データベースの文：「日本の首都は東京である」

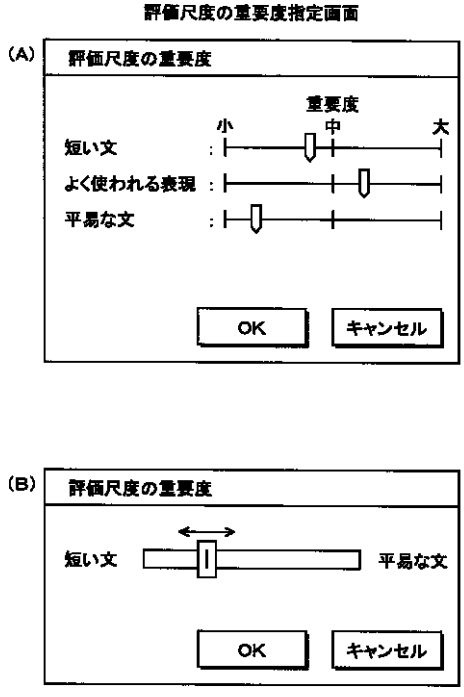
【図７】


言語変換処理の処理フローチャート
START
変換対象文を入力する
S10
変換に必要な変形規則を読み込む
S11
変換対象文を変形規則を用いて変形する
S12
評価尺度の選択画面を表示する
S13
評価尺度の選択情報を入力する
S14
各評価尺度の重要度指定画面を表示する
S15
重要度指定情報を入力する
S16
変形後の表現の各候補を各評価尺度を用いて評価する
S17
各評価尺度の重要度に従って総合的な評価結果を算出する
S18
評価結果の最もよい表現に変形し、変換結果文として出力する
S19
評価尺度を変更するかどうかを問合わせる
S20
評価尺度変更？
S21
YES
NO
END

【図８】


評価尺度選択画面
評価尺度選択
評価尺度を選択してください（複数可）。
短い文
よく使われる表現
平易な文
著名な作家の文体
作家名
夏目 漱石
口語の表現
文章語の表現
二つの文の類似度
選択終了
キャンセル

【図９】


評価尺度の重要度指定画面
(A)
評価尺度の重要度
重要度
小
中
大
短い文
よく使われる表現
平易な文
OK
キャンセル
(B)
評価尺度の重要度
短い文
平易な文
OK
キャンセル

【手続補正書】
【提出日】平成１４年7月２９日（２００２．７．２９）
【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他の表現による文字列に変換するシステムであって，自然言語で記述された変換対象の文字列を入力する入力手段と，前記入力された文字列を所定の変形規則または所定の変形処理プログラムを用いて変形し，変換の候補を生成する変形処理手段と，文字列を変形した結果の表現が目的とする表現になっているかどうかを評価するためのN種類（N≧２）の評価尺度を組み合わせ，各評価尺度に応じて前記変形処理手段により生成された変換の候補を評価し，N種類の評価尺度による総合的な評価結果のよい表現を選択する評価処理手段と，前記選択された表現の変換結果を，目的とする表現に変換された文字列として出力する出力手段とを備え，前記評価処理手段は，変換の候補の評価に用いるN種類の評価尺度を，あらかじめ用意されたM種類（M≧N）の評価尺度の中から外部からの指定により選択する手段を備えることを特徴とする複数尺度の利用による言語変換処理システム。
【請求項２】　前記評価処理手段が変換の候補の評価に用いるN種類の評価尺度に対して，各評価尺度の重要度に関する指定情報を入力する手段を備え，前記評価処理手段は，入力された指定情報に基づいて，個々の評価尺度に対する重要度に応じた評価結果から総合的な評価を行うことを特徴とする請求項１記載の複数尺度の利用による言語変換処理システム。
【請求項３】　前記N種類の評価尺度の少なくとも一つは，変形後の文字列の長短，大量の用例に関する言語データ中に現れる出現頻度もしくは出現確率，所定の平易な文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，口語もしくは文章語で表現された大量の文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，特定の個人の文章集合からなる言語データ中に現れる出現頻度もしくは出現確率，または，変換対象となっている複数の文字列の類似度のいずれかであることを特徴とする請求項１または請求項２記載の複数尺度の利用による言語変換処理システム。
【請求項４】　コンピュータによって，ある自然言語で記述された文字列を他の表現による文字列に変換するためのプログラムであって，自然言語で記述された変換対象の文字列を入力する処理と，前記入力された文字列を所定の変形規則または所定の処理手続きによって変形し，変換の候補を生成する処理と，文字列を変形した結果の表現が目的とする表現になっているかどうかを評価するためのN種類（N≧２）の評価尺度を組み合わせ，各評価尺度に応じて前記変形によって生成された変換の候補を評価し，N種類の評価尺度による総合的な評価結果のよい表現を選択する処理と，前記選択された表現の変換結果を，目的とする表現に変換された文字列として出力する処理と，前記変換の候補の評価に用いるN種類の評価尺度を，あらかじめ用意されたM種類（M≧N）の評価尺度の中から外部からの指定により選択する処理とを，コンピュータに実行させるための複数尺度の利用による言語変換処理プログラム。