

(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2015-138414  
(P2015-138414A)

(43) 公開日 平成27年7月30日(2015.7.30)

(51) Int.Cl.		F I			テーマコード (参考)
<b>G06F 17/27</b>	<b>(2006.01)</b>	G06F 17/27		D	5B091
<b>G06F 17/21</b>	<b>(2006.01)</b>	G06F 17/21	550J		5B109

審査請求 未請求 請求項の数 8 O L (全 25 頁)

(21) 出願番号	特願2014-9862 (P2014-9862)	(71) 出願人	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番1号
(22) 出願日	平成26年1月22日 (2014.1.22)	(74) 代理人	100074099 弁理士 大菅 義之
		(74) 代理人	100133570 弁理士 ▲徳▼永 民雄
		(72) 発明者	角谷 昌剛 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		(72) 発明者	長瀬 友樹 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 機械翻訳装置、翻訳方法、及び、そのプログラム

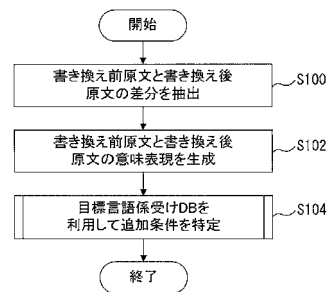
(57) 【要約】

【課題】 翻訳精度を向上させることができる機械翻訳装置、翻訳方法、及び、そのプログラムを提供する。

【解決手段】 機械翻訳装置は、ユーザが入力する入力文を翻訳して訳文を出力する機械翻訳装置である。機械翻訳装置は、ユーザが入力した入力文の書き換え前原文と書き換え後原文から差分を取得し、入力文の意味表現と、差分とから書き換え規則を取得する。そして、機械翻訳装置は、ユーザからの入力文に、取得した書き換え規則を適用して訳文を出力する。

【選択図】 図18

第1の実施形態のフローチャート



**【特許請求の範囲】****【請求項 1】**

ユーザが入力する入力文を翻訳して訳文を出力する機械翻訳装置であって、  
前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得し、  
前記入力文の意味表現と、前記差分とから書き換え規則を取得する規則取得部と、  
前記入力文に、前記規則取得部で取得した前記書き換え規則を適用して前記訳文を出力する翻訳部と、  
を備えることを特徴とする機械翻訳装置。

**【請求項 2】**

翻訳を行ないながらメッセージの送受信を行なっている場合、メッセージの送受信相手からの、前記訳文の意味内容についてのフィードバックを利用して、前記書き換え規則を更新することを特徴とする請求項 1 に記載の機械翻訳装置。

10

**【請求項 3】**

前記意味表現は、入力される文を構成する単語を、それらの修飾関係と意味を含めて、ノードとして表現した形式のデータであることを特徴とする請求項 1 に記載の機械翻訳装置。

**【請求項 4】**

前記書き換え規則は、前記差分に対して、特定の意味と修飾関係を有する前記ノードがあるか否かを判断することによって行なうことを特徴とする請求項 3 に記載の機械翻訳装置。

20

**【請求項 5】**

前記書き換え規則は、置き換え対象の第 1 の単語と、書き換えを行なうべきと判断するために使用する第 2 の単語と、前記第 1 の単語と前記第 2 の単語の修飾関係と、を 1 組として設定したデータであることを特徴とする請求項 1 に記載の機械翻訳装置。

**【請求項 6】**

請求項 1 の機械翻訳装置と、  
2 人以上の利用者がネットワークを介してメッセージの交換を可能とし、前記メッセージは前記機械翻訳装置を用いて翻訳されて送受信され、前記利用者のフィードバックにより前記機械翻訳装置が前記書き換え規則の取得を可能とするサーバと、  
を備える翻訳システム。

30

**【請求項 7】**

コンピュータに、ユーザが入力する入力文を翻訳させて訳文を出力させる翻訳方法であって、  
前記コンピュータは、  
前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得し、  
前記入力文の意味表現と、前記差分とから書き換え規則を取得し、  
前記入力文に、取得した前記書き換え規則を適用して前記訳文を出力する、  
ことを特徴とする翻訳方法。

**【請求項 8】**

コンピュータに、ユーザが入力する入力文を翻訳させて訳文を出力させる、コンピュータ実行可能なプログラムであって、  
前記コンピュータに、  
前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得させ、  
前記入力文の意味表現と、前記差分とから書き換え規則を取得させ、  
前記入力文に、取得した前記書き換え規則を適用して前記訳文を出力させる、  
ことを特徴とするプログラム。

40

**【発明の詳細な説明】****【技術分野】**

50

## 【0001】

本明細書で議論される実施形態は、機械翻訳装置、翻訳方法、及び、そのプログラムに関する。

## 【背景技術】

## 【0002】

自然言語処理システムを用いて、ある言語から別の言語に翻訳することを機械翻訳と呼ぶことがある。機械翻訳装置とは、外国語で書かれた科学技術論文、特許明細書、機器の仕様書や取り扱い説明書、報道記事などを日本文に自動的に翻訳するための装置として知られている。このような機械翻訳の翻訳精度は正解率が7～8割であり、一定の誤訳を含むことが知られている。

10

## 【0003】

例えば、自然言語で記述されたテキストの前編集処理において、テキストの種類の識別と、そのテキストの種類に応じた前編集規則に基づいて検出したテキストの前編集対象部分の単語を標準表記に書き換える技術が知られている。

## 【先行技術文献】

## 【特許文献】

## 【0004】

【特許文献1】特開2000-268034号公報

## 【発明の概要】

## 【発明が解決しようとする課題】

20

## 【0005】

原文を最初に翻訳した訳文には不適切な単語が含まれている。この場合、不適切な単語を置き換える処理を機械翻訳装置が行なうことが考えられる。原文の書き換え規則が単に単語を置き換えるだけのものでは、置き換え処理の誤適用の問題が生じることがある。例えば、原文中に「切る」（動詞）がある場合に、「切る」を「回す」に置き換えることによって、「ボールが見えたらハンドルの少し切る」という文は「ボールが見えたらハンドルの少し回す」と変換することができる。しかし、この規則を「トランプを切る」に適用すると、「トランプを回す」となってしまう。

## 【0006】

そこで、翻訳精度を向上させることができる機械翻訳装置、翻訳方法、及び、そのプログラムを提供することを目的とする。

30

## 【課題を解決するための手段】

## 【0007】

本実施形態の第1の側面における機械翻訳装置は、ユーザが入力する入力文を翻訳して訳文を出力する機械翻訳装置であって、前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得し、前記入力文の意味表現と、前記差分とから書き換え規則を取得する規則取得部と、前記入力文に、前記規則取得部で取得した前記書き換え規則を適用して前記訳文を出力する翻訳部とを備える。

## 【0008】

本実施形態の第2の側面における翻訳方法は、コンピュータに、ユーザが入力する入力文を翻訳させて訳文を出力させる翻訳方法であって、前記コンピュータは、前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得し、前記入力文の意味表現と、前記差分とから書き換え規則を取得し、前記入力文に、取得した前記書き換え規則を適用して前記訳文を出力する。

40

## 【0009】

本実施形態の第3の側面におけるプログラムは、コンピュータに、ユーザが入力する入力文を翻訳させて訳文を出力させる、コンピュータ実行可能なプログラムであって、前記コンピュータに、前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得させ、前記入力文の意味表現と、前記差分とから書き換え規則を取得させ、前記入力文に、取得した前記書き換え規則を適用して前記訳文を出力させる。

50

## 【発明の効果】

## 【0010】

実施形態の機械翻訳装置、翻訳方法、及び、そのプログラムによれば、翻訳精度を向上させることができるという効果を奏する。

## 【図面の簡単な説明】

## 【0011】

【図1】比較例を説明するための図である。

【図2A】比較例の課題であり、事例から自動的に作成した、書き換え規則例を示す図である。

【図2B】比較例の課題であり、誤適用回避のため、人間が規則を追加した例を示す図である。

10

【図3】本発明の実施形態の概要を説明するための図であり、原文解析の結果（意味表現）を示す図である。

【図4】本発明の実施形態の概要を説明するための図であり、意味表現変換規則への追加を示す図である。

【図5A】追加規則生成部を説明するための図である。

【図5B】追加規則生成部を説明するための図である。

【図6】追加対象の規則例を示す図である。

【図7】本発明の概要を説明するための図である。

【図8】本発明の概要を説明するための図である。

20

【図9】第1の実施形態の構成図である。

【図10】第1の実施形態の動作説明図である。

【図11】追加規則要否判定部の処理を説明するための図である。

【図12A】追加規則特定部の処理を説明するための図である。

【図12B】追加規則特定部の処理を説明するための図である。

【図12C】追加規則特定部の処理を説明するための図である。

【図12D】追加規則特定部の処理を説明するための図である。

【図13】書き換え規則が適用された結果翻訳が正解となる例を示す図である。

【図14】誤適用を起こさず翻訳が正解となる例を示す図である。

【図15】書き換え規則のDB格納イメージを示す図である。

30

【図16】書き換え規則のDB格納イメージを示す図である。

【図17】標準的なコンピュータのハードウェア構成の一例を示すブロック図である。

【図18】第1の実施形態のフローチャートである。

【図19】第1の実施形態のフローチャートである。

【図20】第1の実施形態のフローチャートである。

【図21】第1の実施形態のフローチャートである。

【図22】第2の実施形態の構成を説明するための図である。

【図23】第2の実施形態の動作説明図である。

【図24】第2の実施形態のフローチャートである。

【図25】第3の実施形態の動作説明図である。

40

【図26】原文解析の結果を示す図である。

【図27】第3の実施形態の動作説明図である。

【図28】意味表現を説明するための図である。

## 【発明を実施するための形態】

## 【0012】

以下では、機械翻訳をインタラクティブに行う環境で事例を集めることにより、翻訳精度を向上させることができる機械翻訳装置の実施形態について説明する。

## 【0013】

まず、図1～図2Bを参照して、比較例を説明する。次に、図3～図28を参照して、機械翻訳装置の実施形態について説明する。

50

## 【 0 0 1 4 】

< 比較例 >

図 1 は、比較例を説明するための図である。図 1 では、書き換え規則を生成する処理が示されている。

## 【 0 0 1 5 】

図 1 に示すように、書き換え前の文「ハンドルを少し右に切ってください」( S 1 0 ) と、書き換え後の文「ハンドルを少し右に回してください」( S 1 1 ) が用意されているとする。

## 【 0 0 1 6 】

差分解析 ( S 1 2 ) では、書き換え前の文「ハンドルを少し右に切ってください」と、書き換え後の文「ハンドルを少し右に回してください」の差分を求める。その際、S 1 3 に示すように、丁寧語を通常の文に直して「ハンドルを少し右に切る」から「ハンドルを少し右に回す」と書き換えて、両者の差分を取るとする。差分を求めた結果として、「切る」が「回す」に変わっていることが検出される。

10

## 【 0 0 1 7 】

S 1 3 の結果を受けて S 1 4 では書き換えのルール生成を行う。この場合、S 1 5 に示されているように、本文中に「切る」という動詞がある場合、「切る」を「回す」という動詞に置き換えるという規則が自動的に生成される。

## 【 0 0 1 8 】

図 2 A は、比較例の課題であり、事例から自動的に作成した、書き換え規則例を示す図である。

20

## 【 0 0 1 9 】

図 2 A に示すように、原文の書き換え規則は、単純に差分文字列を置き換えるだけでは誤適用の問題が生じる。例えば、動詞を「切る」から「回す」に置き換えるという置き換え規則は、単純に、「切る」が発見されたら、「回す」に置き換えるというものである。したがって、例えば、「トランプを切る」という文が存在した場合、これを単純に「トランプを回す」に置き換えてしまう。

## 【 0 0 2 0 】

図 2 B は、比較例の変形の課題を説明する図である。

図 2 A の比較例の場合、書き換え規則の誤適用が発生する可能性があった。これを改善するために、書き換え規則の適用に人間が適切な追加規則を設定することで、誤適用の問題を回避することができる。

30

## 【 0 0 2 1 】

例えば、人間が与える条件としては、

- 1) 本文中に「切る」( 動詞 ) がある
- 2) 前方に「ハンドル」- 「を」がある

場合に、

「切る」を「回す」( 動詞 ) に置き換える

と言うものが考えられる。

## 【 0 0 2 2 】

この場合、「ハンドルを少し右に切る」は、正しく「ハンドルを少し右に回す」と置き換えられる。また、「トランプを切る」の場合も、この文の中には、「切る」の前に「ハンドル」- 「を」が無い場合、書き換え規則が適用されず、「トランプを切る」という文のままに維持される。

40

## 【 0 0 2 3 】

しかし、人間が追加規則を設定する場合には、追加規則の設定は単語の依存関係についての知識を必要とするため、適切な追加規則が設定できる保証がない。また、上記の方法では、書き換え規則の誤適用回避のための追加規則の設定を自動で設定できない。

## 【 0 0 2 4 】

< 実施形態 >

50

以下に、図3～図21を参照して、第1の実施形態について説明する。

【0025】

本実施形態では、書き換え前/後の原文を構文・意味解析して、単語同士の係受け関係を条件設定に利用する。また、書き換え前/後の、書き換え対象の単語の単語同士の係受けの組み合わせの出現頻度差等を手掛かりに追加規則を設定する。

【0026】

すなわち、翻訳を改善するための原文書き換え事例が入力されたら、

- 1) 書き換え前/後の原文を構文/意味解析してそれぞれの意味表現を生成し、
- 2) 書き換え前/後の文の差分に対応する表現部分を特定し、追加規則の必要性を判断し、
- 3) 目標言語の言語モデル(係受け頻度テーブルなど)を参照して追加規則を特定し、
- 4) 意味表現の変換規則として、書き換え規則データベースを蓄積する。

10

【0027】

ここで、意味表現とは、概念記号同士の関係を有向グラフで表現した構文・意味解析結果である。

【0028】

図3にあるように、書き換え前の原文が「ハンドルを少し右に切ってください」で、書き換え後の原文が「ハンドルを少し右に回してください」だとする。これらを英語の単語に置き換えて、構文/意味解析をした結果が示されている。

【0029】

それぞれ、「ハンドル」、「少し」、「右に」、「切って」が“Steering wheel”、“a little”、“right”、“cut”という英単語に置き換えられている。書き換え後は、「切って」が「回して」になり、“cut”が、“turn”になっている。動詞として、“cut”が検出され、その対象として“Steering wheel”が、方向として“right”が、程度として“a little”が構文/意味解析の結果としての意味表現として得られている。また、書き換え後の文については、動詞として、“cut”ではなく、“turn”が検出されている以外は同様である。

20

【0030】

そして、原文差分に対応する意味表現範囲の妥当性判断を行なう。例えば、差分として検出された単語の1つを1つのノードとした場合、差分に含まれるノードが2つ以上であるか否か判断し、NO(“cut”の部分1つのみ)の場合は、追加規則は必要ないとし、そうでない場合には、追加規則の追加が必要と判断する、などが考えられる。

30

【0031】

意味表現変換規則生成部では、追加規則を含む規則を生成する。意味表現変換規則への追加としては、例えば、差分として検出されたノードに隣接するノード(図3の例では、「切って」に対する「ハンドル」、「少し」、「右に」等)にも条件を追加する事が可能である。

【0032】

例えば、図4に示されるように、差分である“cut”と“turn”に隣接する、意味内容として方向を示すノードがある場合には、“cut”ではなく、“turn”を選択するなどである。

40

【0033】

図5A及び図5Bは、追加規則の取得について説明する図である。

まず、追加規則の取得の準備として、さまざまな例文を書き換え前と書き換え後についてデータベースなどに格納しておく。追加規則生成部は、差分(CUT(切る):TURN(回す))と係受け関係にある単語について、データベース等に格納された例文を検索して、差分に対する当該単語の発生頻度を取得し、別個にデータベースに格納する。なお、例文は、本実施形態の機械翻訳装置が翻訳を行なった文を順次蓄積して、徐々に例文の数を多くしても良い。このようにすることにより、さまざまな例文を蓄積することができ、より正確

50

な追加規則の設定に役立つ。また、例文は、書き換え前と書き換え後それぞれについて、分けて格納しておいても良い。

【0034】

図5Aは、書き換え前の目標言語係受けデータベースの例である。差分である“cut”に対し、目的としての関係を有する修飾語に“Steering wheel”を有する頻度が10と示されている。この数値は、頻度を表すものであれば、どのような単位の値を使ってもよいが、例えば、対応する例文の数などが挙げられる。差分である“cut”に対し、程度としての関係を有する修飾語に“a little”を有する頻度が80であり、方向としての関係を有する修飾語に“right”を有する頻度が3と示されている。

10

【0035】

図5Bは、書き換え後の目標言語係受けデータベースの例である。

差分である“turn”に対し、目的としての関係を有する修飾語に“Steering wheel”を有する頻度が50と示されている。差分である“turn”に対し、程度としての関係を有する修飾語に“a little”を有する頻度が60であり、方向としての関係を有する修飾語に“right”を有する頻度が70と示されている。

【0036】

なお、上記では、例文をデータベースに格納し、頻度を逐次取得して、書き換え前/書き換え後の目標言語係受けデータベースを作成することを説明したが、予め、さまざま単語について、直接、書き換え前/書き換え後の目標言語係受けデータベースを作成してもよい。

20

【0037】

ここで、書き換え前の頻度が小さい単語は、翻訳後の目標言語として不自然であると考えられる。そこで、書き換え前頻度/書き換え後頻度の値で、不自然か否かを判定するようにする。

【0038】

図5A及び図5Bの場合、

<目的>	Steering wheel	:	10 / 50 = 0.2
<程度>	A little	:	80 / 60 = 1.3
<方向>	right	:	3 / 70 = 0.04

30

となるので、方向の頻度値が最少である。したがって、方向を示す“right”を追加規則に設定する。すなわち、差分である“cut”に対し、方向としての関係を有する修飾語に“right”を有する場合には、“cut”を“turn”に置き換えるようにする。

【0039】

書き換え前頻度/書き換え後頻度の値が閾値を下回った単語を別々に追加規則設定した規則を追加することも可能である。

【0040】

閾値を0.3とすると、

<目的>	Steering wheel	:	10 / 50 = 0.2	閾値以下
<程度>	A little	:	80 / 60 = 1.3	
<方向>	right	:	3 / 70 = 0.04	閾値以下

40

よって、「cut~right”から“turn~right”という追加規則に加えて、図6に示されるような条件も追加する。図6の追加規則は、差分である“cut”に対し、目的としての関係を有する修飾語に“Steering wheel”を有する場合には、“cut”を“turn”に置き換えるようにする、というものである。

【0041】

以上のような追加規則を追加した後に、日英翻訳を行なう例が図7である。

入力文として、「彼はハンドルを素早く切る」という文が与えられたとする。原文解析の結果の意味表現は、(a)のようになる。主語として“he”が、対象として“Ste

50

ering wheel”が、様態として“quickly”が検出される。ここで、動詞である“cut”に対し、目的としての関係を有する修飾語に“Steering wheel”が発見される。したがって、上記追加規則から、意味表現変換として、“cut”を“turn”に置き換える。結果として得られる日英翻訳文は、「He turns the steering wheel quickly.」となる。

【0042】

また、図8のような日英翻訳の例も可能である。

すなわち、入力文が「ハンドルを持ってねじ山を切ってください」である場合、原文解析の結果の意味表現は、図8のようになる。すなわち、動詞“cut”に対し、対象が“Screw thread”であり、付帯語として、“possess”、この付帯語の対象として“Steering wheel”が検出される。したがって、“cut”を修飾する語句として方向としての関係を有する“right”あるいは、目的としての関係を有する“Steering wheel”が存在しないので、置き換えのための規則に対応するものはない。したがって、“cut”は置き換えられずにそのまま維持され、日英翻訳文として、「Please cut the screw thread with the steering wheel.」が得られる。

【0043】

図9は、本実施形態の機械翻訳装置のブロック構成図である。

機械翻訳装置100は、例文として格納されている書き換え前原文202、書き換え後原文204を格納するデータベースに接続されている。DB規則取得部101は、書き換え前原文202と書き換え後原文204を読み込み、追加規則を取得する。まず、差分抽出部102は、書き換え前原文202と書き換え後原文204の差分を検出する。検出された差分は、意味表現生成部110からの、翻訳対象の原文200の文構造とともに、追加規則要否判定部106に入力され、追加規則の設定の要否が判定される。追加規則が必要と判定された場合には、差分のデータと共に、追加規則を設定すべき旨が追加規則特定部107に入力される。追加規則特定部107は、前述の目標言語係受けデータベース(DB)104を参照して追加規則を決定し、書き換え規則データベース(DB)108に登録する。

【0044】

翻訳対象の原文200は、機械翻訳部103の意味表現生成部110に入力される。意味表現生成部110では、原文200が翻訳目標の言語の単語に置き換えられ、文構造が単語単位で解析され、原文が動詞とそれを修飾する単語などのノードとして検出される。意味表現生成部110で意味表現が得られると、原文200の意味表現は意味表現置換部112に入力される。意味表現置換部112は、書き換え規則DB108を参照して、原文の一部の単語の置き換えを行う。そして、訳文生成部114は、訳を生成し、訳文206として出力する。

【0045】

図10～図14を参照して、本実施形態の機械翻訳装置の動作を説明する。

まず、書き換え規則の取得を行なう。書き換え前の原文が「ハンドルを少し右に切ってください」であり、書き換え後の原文が「ハンドルを少し右に回してください」である場合、それぞれについて、原文解析を行なって、図10のような意味表現を得る。

【0046】

(a)が書き換え前の原文の意味表現であり、(b)が書き換え後の原文の意味表現である。両者の差分に当たるのが“cut”300と“turn”400である。これらに対し、対象としての修飾語が“Steering wheel”302、402である。方向としての修飾語が“right”304、404である。程度としての修飾語が“alittle”306、406である。

【0047】

次に、追加規則要否判定部の処理として、原文差分に対応する意味表現範囲の妥当性の判断を行なう。図11の例では、原文差分は、動詞の“cut”と“turn”のみであ

10

20

30

40

50



る。“cut”を全て“turn”に書き換えるような条件の規則を追加すると、規則の過適用が増えて、翻訳結果がかえって悪くなる可能性がある。そこで、条件に、差分となっている一定数のノード（閾値以上のノード数）がないと条件の追加を認めない事することで、規則の誤適用を防ぐ。例えば、差分となるノードが2つ以上の場合（閾値=2）、追加規則を必要とする、など、隣接するノードも条件に加え、「～を切る」、「～で切る」などのような場合に規則を適用する、などにする。

【0048】

図12A～図12Dは、追加規則特定部の処理を説明する図である。

目標言語係受けデータベースを利用して、差分のノードと係受け関係にある単語の中から追加規則に含める単語を選択する。

10

【0049】

図12Aは、書き換え前の目標言語係受けデータベースの例であり、図12Bは、書き換え後の目標言語係受けデータベースの例である。それぞれの差分単語に対する係受け関係にある単語の、書き換え前頻度/書き換え後頻度の値を計算する。すると、図12A及び図12Bから分かるように、“cut”と“right”の組み合わせが書き換え前に最小頻度である事が分かる。また、“cut”と“right”の組み合わせの頻度の、“turn”と“right”の組み合わせの頻度に対する比の値が最小であることが理解される。これは、“cut”と“right”の組み合わせが不自然であることを意味していると考えられる。したがって、追加規則に含める単語として“right”を選択する。

20

【0050】

すなわち、図12Cにあるように、書き換え前の“cut”に対し、方向を表す修飾語として、“right”があった場合、“cut”を“turn”に置き換える処理を行なう。

【0051】

また、書き換え前頻度/書き換え後頻度の値が所定の閾値を下回った単語を別々に追加規則として追加設定してもよい。例えば、閾値を0.3とすると、

<目的> Steering wheel : 10/50 = 0.2 閾値以下

<程度> A little : 80/60 = 1.3

<方向> right : 3/70 = 0.04 閾値以下

30

であるが、この場合、“right”の他に、“Steering wheel”が追加規則に含める単語として挙げられる。

【0052】

この場合、図12Dにあるように、書き換え前の“cut”に対し、対象を表す修飾語として、“Steering wheel”があった場合、“cut”を“turn”に置き換える処理を行なう。

【0053】

なお、ここでは、閾値を0.3としたが、この閾値は、本実施形態の機械翻訳装置を設計・製造する者が、実験や経験などに基づいて、任意に決定してよい。

【0054】

40

図13及び図14は、翻訳動作を説明する図である。

図13において、入力文として「彼はハンドルを素早く切る」が与えられると、最初の翻訳として、“He cuts Steering wheel quickly”が得られたとする。これを原文解析し、意味表現を得る。すると、動詞“cut”に対し、主語として“he”、対象として“Steering wheel”、様態として“quickly”がノードとして得られる。

【0055】

次に、書き換え規則を適用する。今の場合、書き換え前の“cut”に対し、対象を表す修飾語として、“Steering wheel”がある場合、“cut”を“turn”に置き換える処理を行なう、という追加規則が適用される。したがって、意味表現変

50

換の結果として、“cut”の代わりに“turn”が採用される。そして、この意味表現変換の結果を用いて訳文を生成すると、“He turns the steering wheel quickly.”という訳文が得られる。

【0056】

図14においては、入力文として「ハンドルを持ってねじ山を切ってください」が与えられると、最初の翻訳として、“Please cut the screw thread with the steering wheel.”が得られたとする。これを原文解析して、意味表現を得る。すると、動詞“cut”に対し、対象として“Screw thread”、付帯語として“possess”、付帯語の対象として“Steering wheel”が得られる。

10

【0057】

次に、書き換え規則を適用するが、今の場合、対応する規則が無いので、意味表現変換においては、単語の置き換えは行なわれない。したがって、結果として得られる訳文は、“Please cut the screw thread with the steering wheel.”である。

【0058】

図15及び図16は、書き換え規則データベースの格納データの例を示す図である。

図15の格納イメージにあるように、書き換え規則データベース(DB)108に格納されるデータは、書き換え規則ごとに格納され、各書き換え規則のデータには、書き換え前のノードを表すデータと、書き換え後のノードを表すデータが格納される。

20

【0059】

図16のより具体的なイメージでは、書き換え規則ごとに、ノードのデータをタプル(tuple)として格納する。タプルは、意味表現に含まれる1つのエッジ(ノード間がなすつながり)を表す「Fromノード」、「アーク」、「Toノード」の3項の組である。意味表現全体は、複数のタプルで表す事が可能である。アークとは、FromノードからToノードへの意味的關係である。例えば、“cut”に対して“right”は、方向を意味する修飾語であるが、この場合、アークは「方向」となる。

【0060】

図16においては、レコード番号が1と2の書き換え規則が例示されている。レコード番号が1のレコードには、最初のレコードに、格納されるタプル数が記述される。また、レコード1には、書き換え前のノードと書き換え後のノードを記述している。ここで、各タプルは、「Fromノード」、「アーク」、「Toノード」からなり、書き換え前のノードのデータをタプル[0]に、書き換え後のノードのデータをタプル[1]に格納するとしている。いずれの場合も、1つの書き換え規則に属するので、レコード番号は、1となっている。書き換え前のノードデータであるタプル[0]の値は、(cut、方向、right)であり、書き換え後のノードデータであるタプル[1]の値は、(turn、方向、right)である。

30

【0061】

また、別の書き換え規則は、レコード番号2のレコード2として記述されている。レコード2には、同様に、タプル数と、書き換え前と書き換え後のノードのデータが記述されている。レコード2の書き換え前のタプル[2]の値は、(cut、対象、steering wheel)で、書き換え後タプル[3]の値は、(turn、対象、steering wheel)である。

40

【0062】

図17は、本実施形態の機械翻訳装置をコンピュータで実現する場合のハードウェア環境図である。

【0063】

本実施形態の機械翻訳装置を実現するコンピュータ500は、CPU508によって制御される。CPU508は、バス510を介して、ROM506、RAM504、ハードディスク装置502、入力装置518、表示装置516、インタフェース装置514、記

50

録媒体駆動装置 5 1 2 に接続される。

【 0 0 6 4 】

R O M 5 0 6 は、B I O S など、コンピュータ 5 0 0 を動作させるのに基本的な基本プログラムを格納し、C P U 5 0 8 が、基本プログラムを実行することにより、コンピュータ 5 0 0 の入出力等を可能とする。

【 0 0 6 5 】

R A M 5 0 4 は、本実施形態の機械翻訳を実行するプログラム等を内部に展開し、C P U 5 0 8 が当該プログラムを実行可能とする。

【 0 0 6 6 】

ハードディスク装置 5 0 2 は、R A M 5 0 4 に展開されるべきプログラムや、当該プログラムを実行するのに必要なデータ等を格納する。ハードディスク装置 5 0 2 は、本実施形態の機械翻訳を行なうプログラムを格納することができ、目標言語係受け D B や、書き換え規則 D B を内部に構成しても良い。

10

【 0 0 6 7 】

入力装置 5 1 8 は、キーボードやマウスなどであり、ユーザからコンピュータ 5 0 0 に情報を入力するために使用される。本実施形態の機械翻訳装置の場合、ユーザは、入力装置 5 1 8 を使用して、翻訳したい文をコンピュータ 5 0 0 に入力することができる。

【 0 0 6 8 】

表示装置 5 1 6 は、C R T や液晶ディスプレイなどであり、入力装置 5 1 8 から入力された情報や、C P U 5 0 8 が演算した結果の翻訳文等をユーザに提示するために用いられる。

20

【 0 0 6 9 】

記録媒体駆動装置 5 1 2 は、C D、D V D、B l u - r a y (登録商標)、フレキシブルディスク、I C メモリ等の可搬型記録媒体 5 2 0 からデータを読取ったり、可搬型記録媒体 5 2 0 にデータを記録したりする。可搬型記録媒体 5 2 0 は、ハードディスク装置 5 0 2 と同様で、R A M 5 0 4 に展開されるべきプログラムや、当該プログラムの実行に必要なデータ等を格納する。

【 0 0 7 0 】

インタフェース装置 5 1 4 は、コンピュータ 5 0 0 をネットワーク(不図示)を介して、他のコンピュータと接続する。例えば、ユーザは、他のコンピュータ上に構築された目標言語係受け D B や、書き換え規則 D B を使用して、コンピュータ 5 0 0 上で、本実施形態の機械翻訳プログラムを実行しても良い。また、ネットワークを介して、他のコンピュータから入力された文を、コンピュータ 5 0 0 が翻訳処理しても良い。

30

【 0 0 7 1 】

図 1 8 ~ 図 2 1 は、第 1 の実施形態の動作を説明するフローチャートである。

図 1 8 ~ 図 2 0 は、書き換え規則の取得処理を示すフローチャートである。

【 0 0 7 2 】

図 1 8 において、処理を開始すると、ステップ S 1 0 0 において、D B 規則取得部は、書き換え前原文と書き換え後原文の差分を抽出する。次に、ステップ S 1 0 2 において、D B 規則取得部は、書き換え前原文と書き換え後原文の意味表現を生成する。そして、ステップ S 1 0 4 において、D B 規則取得部は、目標言語係受け D B を利用して追加規則を特定した後、処理を終了する。

40

【 0 0 7 3 】

図 1 9 は、図 1 8 のステップ S 1 0 4 のサブルーチンの処理を示すフローチャートである。

【 0 0 7 4 】

ステップ S 2 0 2 において、D B 規則取得部は、書き換え前原文と書き換え後原文の差分に対応する意味表現を特定する。ステップ S 2 0 4 において、D B 規則取得部は、差分に対応する意味表現に含まれるノード数が閾値以上か否かを判断する。上記説明では、閾値は 2 としていた。ステップ S 2 0 4 の判断で、当該ノード数が閾値より小さいと判断さ

50

れた場合には ( N o )、処理を終了する。ステップ S 2 0 4 の判断で、当該ノード数が閾値以上であると判断された場合には ( Y e s )、ステップ S 2 0 6 において、追加規則の特定が行なわれた後、処理を終了する。

【 0 0 7 5 】

図 2 0 は、図 1 9 のステップ S 2 0 6 のサブルーチンの処理を示すフローチャートである。

【 0 0 7 6 】

ステップ S 3 0 2 において、D B 規則取得部は、差分に対応する意味表現に未処理のノードがあるか否かを判断する。ステップ S 3 0 2 の判断で、未処理のノードがないと判断された場合には ( N o )、D B 規則取得部は、そのまま処理を終了する。ステップ S 3 0 2 の判断で、未処理のノードがあると判断された場合には ( Y e s )、ステップ S 3 0 4 において、D B 規則取得部は、未処理のノードから処理対象のノードを 1 つ選択する。ステップ S 3 0 6 において、D B 規則取得部は、処理対象のノードに隣接する別のノードのうち、以下の条件を満たすノードが存在するか否かを判断する。

10

・書き換え規則に未追加、かつ

・目標言語係受け D B の「書き換え前頻度 / 書き換え後頻度」の値が閾値以下。

【 0 0 7 7 】

ステップ S 3 0 6 の判断において、上記条件を満たすノードが存在しないと判断された場合には ( N o )、D B 規則取得部は、ステップ S 3 0 2 に戻って、処理を繰り返す。ステップ S 3 0 6 の判断において、上記条件を満たすノードが存在すると判断された場合には ( Y e s )、D B 規則取得部は、ステップ S 3 0 8 において、ノードを書き換え規則に追加した後、ステップ S 3 0 6 に戻って、処理を繰り返す。

20

【 0 0 7 8 】

図 2 1 は、機械翻訳部の翻訳処理のフローチャートである。

翻訳対象の原文が入力されると、ステップ S 4 0 2 において、機械翻訳部は、原文から意味表現を生成する。ステップ S 4 0 4 において、機械翻訳部は、書き換え規則 D B を参照し、意味表現が一致する書き換え規則を適用する。ステップ S 4 0 6 において、機械翻訳部は、意味表現から訳文を生成し、出力して処理を終了する。

【 0 0 7 9 】

図 2 2 ~ 図 2 4 は、第 2 の実施形態を説明する図である。

30

図 2 2 にあるように、機械翻訳を利用した外国人とのチャットの中で、相手が理解できなくて、相手から聞き返された場合、元の発信文を別の表現に直すと理解されるケースがある。このとき、表現を修正する前と後の発信文から書き換え規則を取得することが考えられる。

【 0 0 8 0 】

図 2 2 において、利用者 2 が自分であり、利用者 1 が外国人だとする。両者は、チャット・サーバーを使ってチャットするが、母国語が異なるので翻訳エンジンで自分の母国語を相手の母国語に機械翻訳しながらチャットをする。そして、相手からチャットの内容の意味が分からない旨の返信を受けた場合、元のチャット内容を修正して相手に送る。それによって、相手が自分のチャット内容を理解できた場合、修正前と修正後のチャットの原文を書き換え規則登録のために、本実施形態の機械翻訳装置に送信する。当該機械翻訳装置は、送信されてきた修正前と修正後の原文から、前述した方法で、新しい書き換え規則を生成し、格納する。当該機械翻訳装置の翻訳エンジンを使ったチャットでは、翻訳エンジンは、以降、新しい書き換え規則を用いて翻訳を行なうので、チャットにおいて同じように相手に理解されないという場合を削減することができる。

40

【 0 0 8 1 】

図 2 3 は、機械翻訳を利用したチャットの端末画面のイメージ図である。

利用者 2 から利用者 1 に、「ハンドルを少し右に切ってください」という原文を送ろうとしたとき、翻訳エンジンが、「 P l e a s e c u t t h e s t e e r i n g w h e e l r i g h t a l i t t l e 」と翻訳して、利用者 1 に提示したとする。す

50

ると、この訳文は間違っているので、利用者1から利用者2には、“Can't understand”など、利用者1が利用者2のチャット内容を理解できない旨が提示される。そこで、利用者2は、書き換え前原文を修正して、書き換え後原文「ハンドルを少し右に回してください」を入力したとする。翻訳エンジンは、これを翻訳して、“Please turn the steering wheel right a little”を利用者1に提示する。

**【0082】**

この文で、利用者1がチャットの意味を理解した場合、利用者2は、書き換え後原文の横に表示されている「追加規則取得」ボタンを押す（クリックする）。すると、書き換え後原文の1つ前の原文が書き換え前原文として取得され、機械翻訳装置に送信される。機械翻訳装置は、送られてきた書き換え前と書き換え後の原文から書き換え規則を生成し、以後のチャットの翻訳に利用する。すると、以後のチャットにおいて、利用者2から利用者1に「ハンドルを切ってください」という原文を提示したとすると、利用者1に提示される翻訳文は、“Please turn the steering wheel”となる。

10

**【0083】**

図24は、第2の実施形態の規則取得処理を示すフローチャートである。

チャットが行なわれている最中において、ステップS502では、チャットの利用者が追加規則取得ボタンをクリックしたか否かが判断される。ステップS502の判断で、追加規則取得ボタンがクリックされていないと判断された場合は（No）、チャット用ソフトウェアは処理を終了する。ステップS502の判断で、追加規則取得ボタンがクリックされたと判断された場合には（Yes）、ステップS504で、チャット用ソフトウェアは、追加規則取得ボタンが押された原文と、1つ前の原文を、それぞれ、「書き換え後原文」、「書き換え前原文」として翻訳サーバ（本実施形態の機械翻訳装置）に送信する。ステップS506において、翻訳サーバは追加規則取得処理を実施し、処理を終了する。ステップS506の追加規則取得処理は、図18～図20のフローチャートで説明した処理と同様である。

20

**【0084】**

図25～図27は、第3の実施形態を説明する図である。

第3の実施形態は、本実施形態の機械翻訳装置を英日翻訳に適用した場合である。

30

**【0085】**

図25に示されるように、書き換え前原文が“The noise is Issued by the transistor”で、書き換え後原文が“The noise is generated by the transistor”であるとする。書き換え前原文の日本語訳は「雑音はトランジスタによって発行される」であり、誤った日本語となっているが、書き換え後原文の日本語訳は「雑音はトランジスタによって発生する」となって、正しい日本語となっている。

**【0086】**

機械翻訳装置は、これらを原文解析し、意味表現を得る。図26にあるように、書き換え前の原文の意味表現においては、“issue”を動詞として、対象としての修飾語である“noise”と、手段としての修飾語である“transistor”がノードとして得られる。書き換え後の原文の意味表現においては、“generate”を動詞として、対象としての修飾語である“noise”と、手段としての修飾語である“transistor”がノードとして得られる。ここで、両者の差分は、“issue”と“generate”である。

40

**【0087】**

次に、書き換え規則を設定するが、ここでは仮に、対象としての修飾語に“noise”を持つ場合、“issue”を“generate”に置き換える、という規則が得られたとする。

**【0088】**

50

図27にあるように、入力文として、“We found out that the noise was issued by the power supply cable”という文が入力されたとする。書き換え規則を適用しないで翻訳を行なった場合の訳文は、「私たちは、雑音が電力ケーブルによって発行されたのを見つけました」となり、誤った日本語となる。

【0089】

機械翻訳装置は、この入力文を原文解析し、結果として、意味表現を得る。ここでは、動詞としての“issue”に対し、対象としての修飾語“noise”と、手段としての修飾語“cable”と、“cable”の限定語としての“electric power”がノードとして得られる。

10

【0090】

ここに書き換え規則を適用する。上記の例では、動詞としての“issue”に対し、対象としての修飾語“noise”が存在するので、“issue”を“generate”に置き換える、という規則に合致する。したがって、機械翻訳装置は、“issue”を“generate”に置き換える意味表現変換を行なう。

【0091】

以上の結果、英日翻訳の出力訳文として、「私たちは、雑音が電力ケーブルで発生したのを見つけました」という文が得られる。

【0092】

図28は、構文・意味解析結果である意味表現を用いた翻訳処理の概念を説明する図である。

20

【0093】

原文として「太郎は秋葉原で富士通のパソコンと携帯電話を買った。」という文が入力されると、翻訳辞書を参照して、機械翻訳装置は形態素解析を行なう。文を形態素に分割した結果、原文は、「太郎」、「秋葉原」、「富士通」、「パソコン」、「携帯電話」、「買」、「っ」、「た」に分割される。その後、構文解析が行なわれ、係受け構造が形成される。係受け構造を意味解析し、原文の概念構造を構築する。ここでは、“buy”が動詞であり、“Taro”が行為者、“Akihabara”が場所、動詞の対象が“cell-phone”であり、これに関連した対象が“PC”と“Fujitsu”であることが理解される。

30

【0094】

そして、翻訳辞書を参照して、概念構造から文を生成し、生成された文が訳文として出力される。ここでは、訳文として“Taro bought the personal computer and the cellular phone of Fujitsu in Akihabara”という文が得られている。

【0095】

以上の構成によると、書き換え前頻度/書き換え後頻度のデータベースを用いることにより、訳語の誤適用回避のための規則を自動で設定することが可能となる。また、チャット翻訳システムで、訳文の意味が伝わらないときに、送り手が意味を変えずに別表現を送信する過程から、翻訳失敗文と成功文を蓄積することで、自律成長型のサービスを実現することが可能となる。

40

【0096】

以上、本発明の実施形態について説明したが、本発明は上述した実施形態に限定されるものではなく、他の様々な変更が可能である。

【0097】

なお、以上までに説明した実施形態に関し、更に以下の付記を開示する。

(付記1)

ユーザが入力する入力文を翻訳して訳文を出力する機械翻訳装置であって、前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得し、前記入力文の意味表現と、前記差分とから書き換え規則を取得する規則取得部と、

50

前記入力文に、前記規則取得部で取得した前記書き換え規則を適用して前記訳文を出力する翻訳部と、  
を備えることを特徴とする機械翻訳装置。

(付記 2)

翻訳を行ないながらメッセージの送受信を行なっている場合、メッセージの送受信相手からの、前記訳文の意味内容についてのフィードバックを利用して、前記書き換え規則を更新することを特徴とする付記 1 に記載の機械翻訳装置。

(付記 3)

前記意味表現は、入力される文を構成する単語を、それらの修飾関係と意味を含めて、ノードとして表現した形式のデータであることを特徴とする付記 1 に記載の機械翻訳装置。

10

(付記 4)

前記書き換え規則は、前記差分に対して、特定の意味と修飾関係を有する前記ノードがあるか否かを判断することによって行なうことを特徴とする付記 3 に記載の機械翻訳装置。

(付記 5)

前記書き換え規則は、置き換え対象の第 1 の単語と、書き換えを行なうべきと判断するために使用する第 2 の単語と、前記第 1 の単語と前記第 2 の単語の修飾関係と、を 1 組として設定したデータであることを特徴とする付記 1 に記載の機械翻訳装置。

(付記 6)

日英翻訳を行なうことを特徴とする付記 1 に記載の機械翻訳装置。

20

(付記 7)

英日翻訳を行なうことを特徴とする付記 1 に記載の機械翻訳装置。

(付記 8)

付記 1 の機械翻訳装置と、

2 人の利用者がネットワークを介してメッセージの交換を可能とし、前記メッセージは前記機械翻訳装置を用いて翻訳されて送受信され、前記利用者のフィードバックにより前記機械翻訳装置が前記書き換え規則の取得を可能とするサーバと、  
を備える翻訳システム。

(付記 9)

コンピュータに、ユーザが入力する入力文を翻訳させて訳文を出力させる翻訳方法であって、

30

前記コンピュータは、

前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得し、

前記入力文の意味表現と、前記差分とから書き換え規則を取得し、

前記入力文に、取得した前記書き換え規則を適用して前記訳文を出力する、  
ことを特徴とする翻訳方法。

(付記 10)

コンピュータに、ユーザが入力する入力文を翻訳させて訳文を出力させる、コンピュータ実行可能なプログラムであって、

40

前記コンピュータに、

前記ユーザが入力した前記入力文の書き換え前原文と書き換え後原文から差分を取得させ、

前記入力文の意味表現と、前記差分とから書き換え規則を取得させ、

前記入力文に、取得した前記書き換え規則を適用して前記訳文を出力させる、  
ことを特徴とするプログラム。

【符号の説明】

【0098】

100 機械翻訳装置

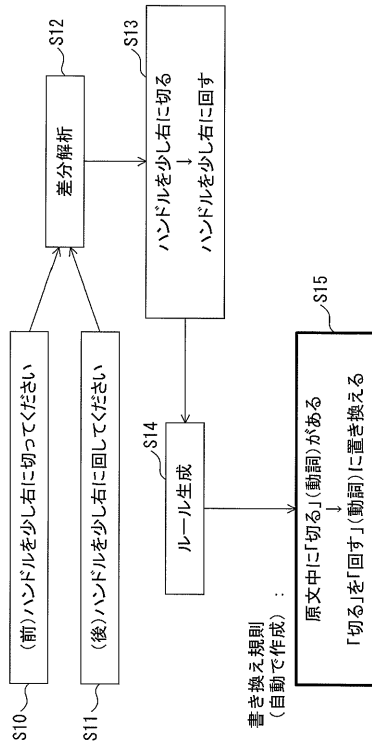
50

1 0 1	D B 規則取得部		
1 0 2	差分抽出部		
1 0 3	機械翻訳部		
1 0 4	目標言語係受け D B		
1 0 6	追加規則要否判定部		
1 0 7	追加規則特定部		
1 0 8	書き換え規則 D B		
1 1 0	意味表現生成部		
1 1 2	意味表現置換部		
1 1 4	訳文生成部		10
2 0 0	原文		
2 0 2	書き換え前原文		
2 0 4	書き換え後原文		
2 0 6	訳文		
3 0 0 ~ 3 0 6、4 0 0 ~ 4 0 6		ノード	
5 0 0	コンピュータ		
5 0 2	ハードディスク装置		
5 0 4	R A M		
5 0 6	R O M		
5 0 8	C P U		20
5 1 0	バス		
5 1 2	記録媒体駆動装置		
5 1 4	インタフェース装置		
5 1 6	表示装置		
5 1 8	入力装置		
5 2 0	可搬型記録媒体		



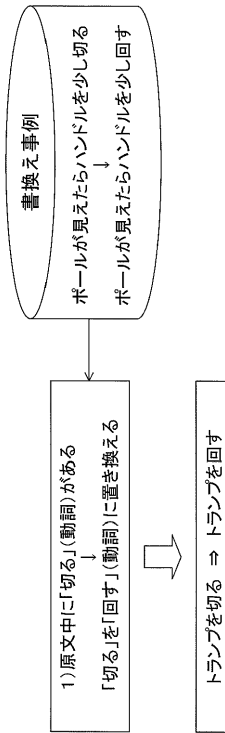
【 図 1 】

比較例を説明するための図



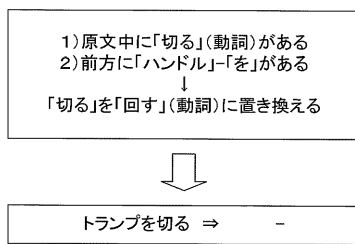
【 図 2 A 】

比較例の課題であり、事例から自動的に作成した、書き換え規則例を示す図



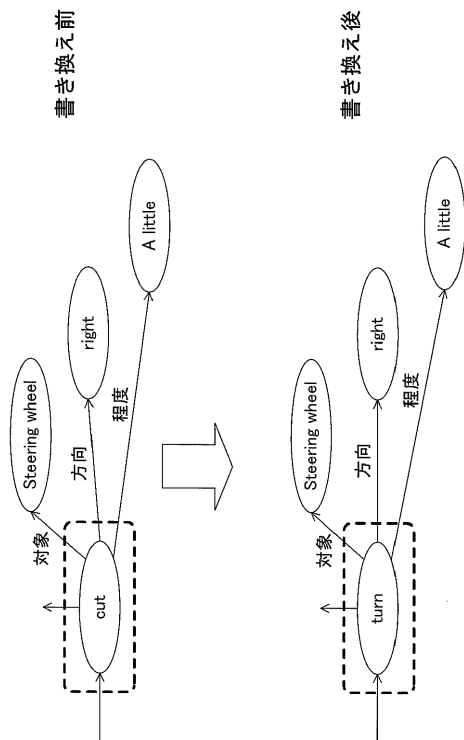
【 図 2 B 】

比較例の課題であり、誤適用回避のため、人間が規則を追加した例を示す図



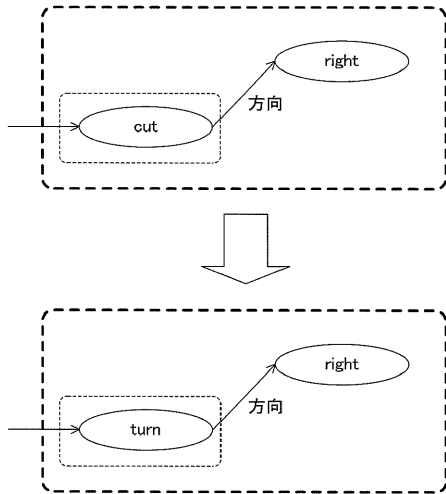
【 図 3 】

本発明の実施形態の概要を説明するための図であり、原文解析の結果(意味表現)を示す図



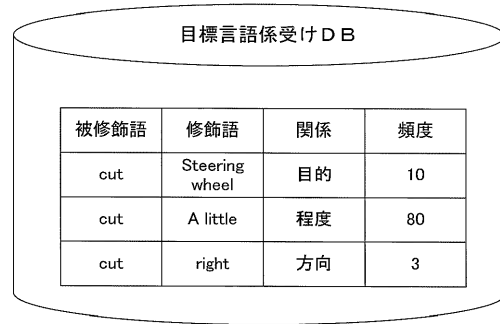
【 図 4 】

本発明の実施形態の概要を説明するための図であり、意味表現変換規則への追加を示す図



【 図 5 A 】

追加規則生成部を説明するための図



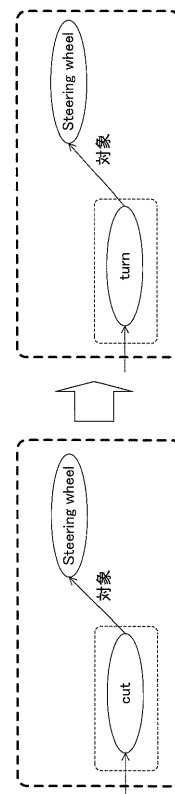
【 図 5 B 】

追加規則生成部を説明するための図



【 図 6 】

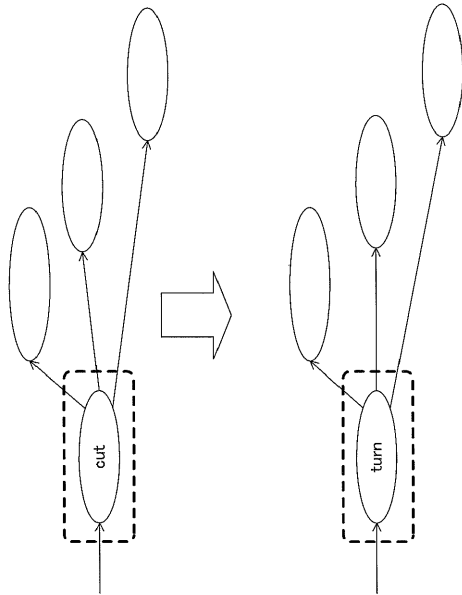
追加対象の規則例を示す図





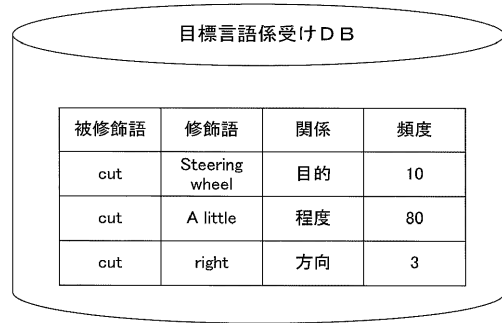
【 図 1 1 】

追加規則要否判定部の処理を説明するための図



【 図 1 2 A 】

追加規則特定部の処理を説明するための図



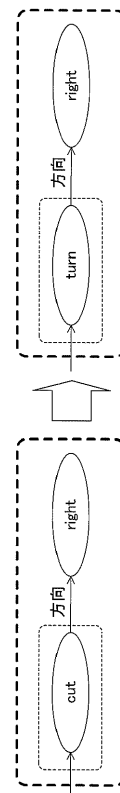
【 図 1 2 B 】

追加規則特定部の処理を説明するための図



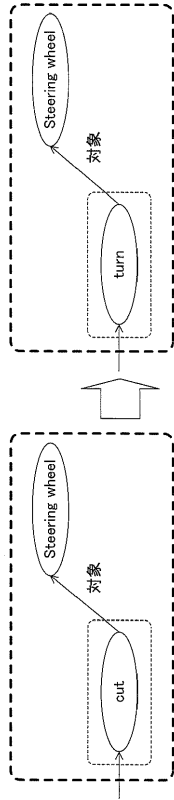
【 図 1 2 C 】

追加規則特定部の処理を説明するための図



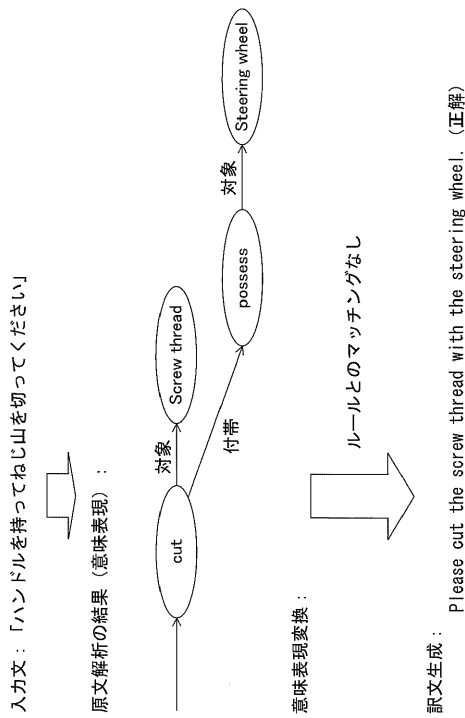
【 図 1 2 D 】

追加規則特定部の処理を説明するための図



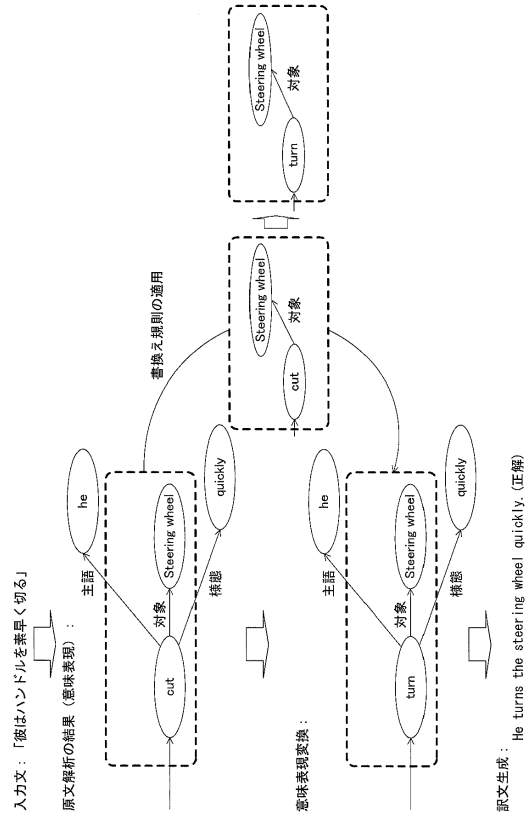
【 図 1 4 】

誤適用を起こさず翻訳が正解となる例を示す図



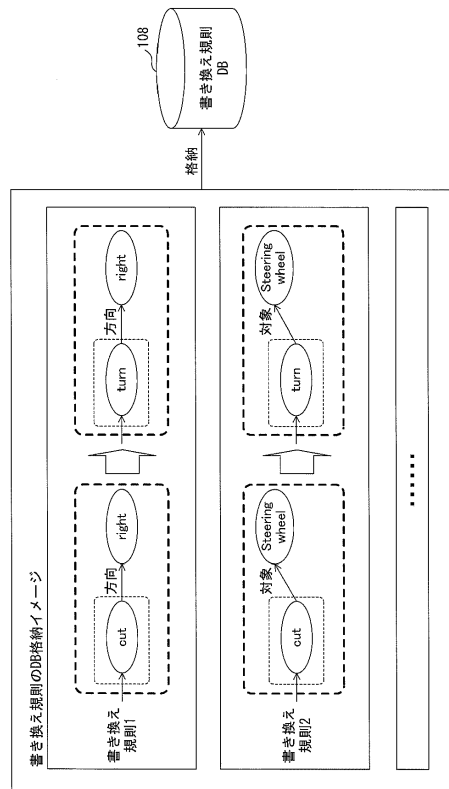
【 図 1 3 】

書き換え規則が適用された結果翻訳が正解となる例を示す図



【 図 1 5 】

書き換え規則のDB格納イメージを示す図



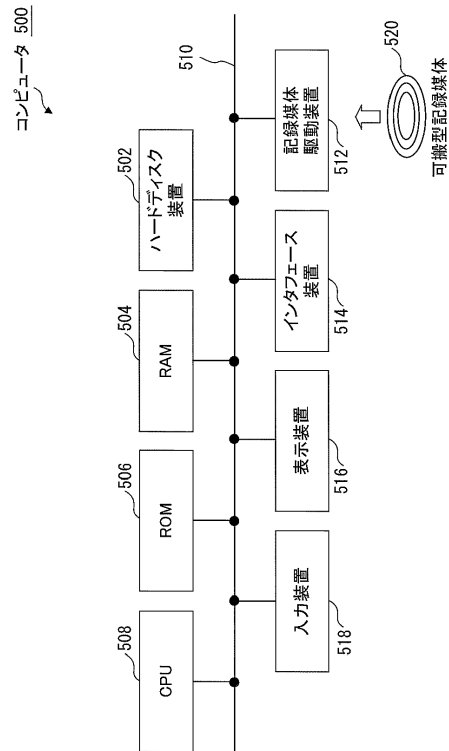
【図16】

書き換え規則のDB格納イメージを示す図

レコード#	項目	値
1	タプル数	2
1	書換え前 タプル[0]	(cut, 方向, right)
1	書換え後 タプル[1]	(turn, 方向, right)
2	タプル数	2
2	書換え前 タプル[2]	(cut, 対象, steering wheel)
2	書換え後 タプル[3]	(turn, 対象, steering wheel)
3	...	...

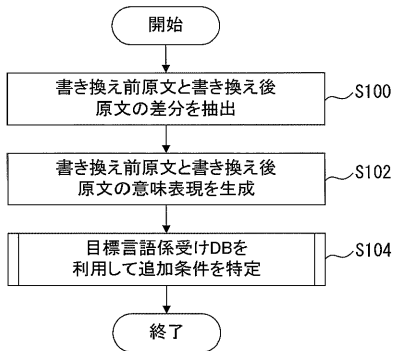
【図17】

標準的なコンピュータの  
ハードウェア構成の一例を示すブロック図



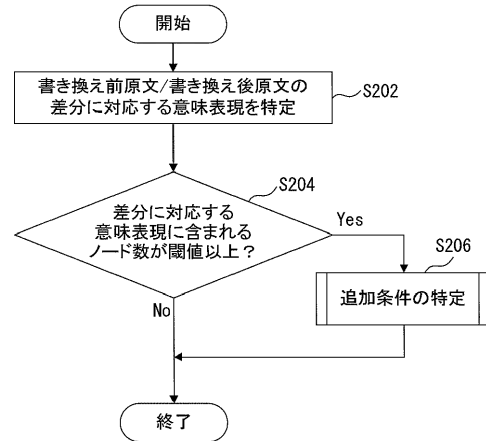
【図18】

第1の実施形態のフローチャート



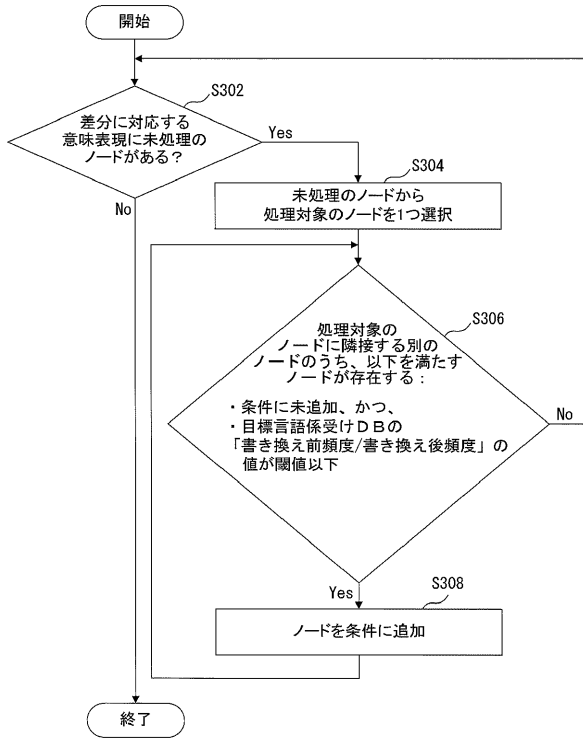
【図19】

第1の実施形態のフローチャート



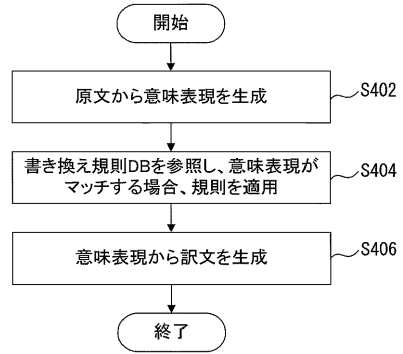
【 図 2 0 】

第1の実施形態のフローチャート



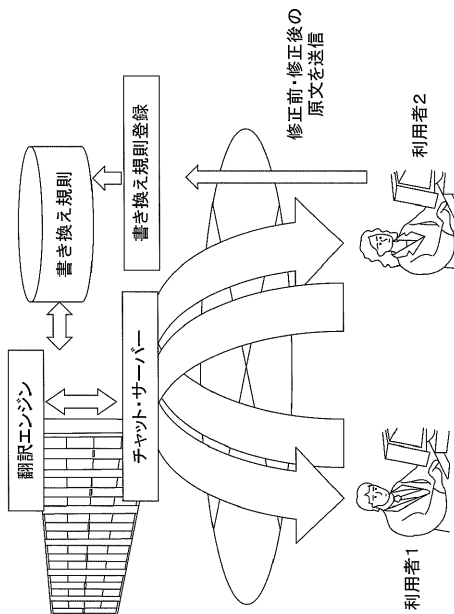
【 図 2 1 】

第1の実施形態のフローチャート



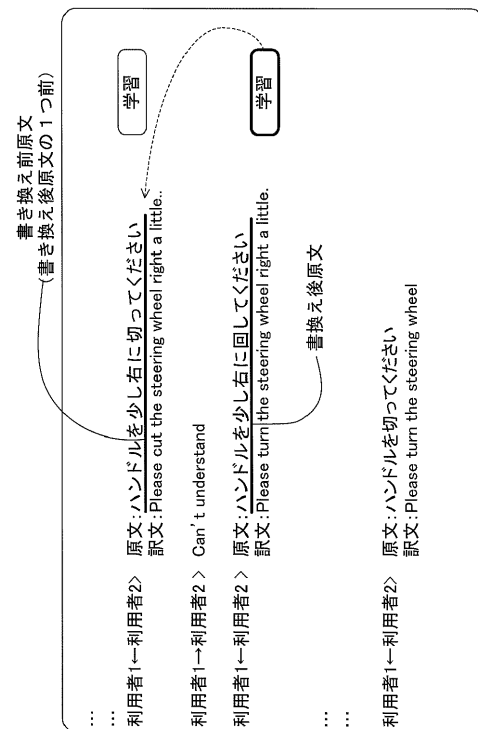
【 図 2 2 】

第2の実施形態の構成を説明するための図



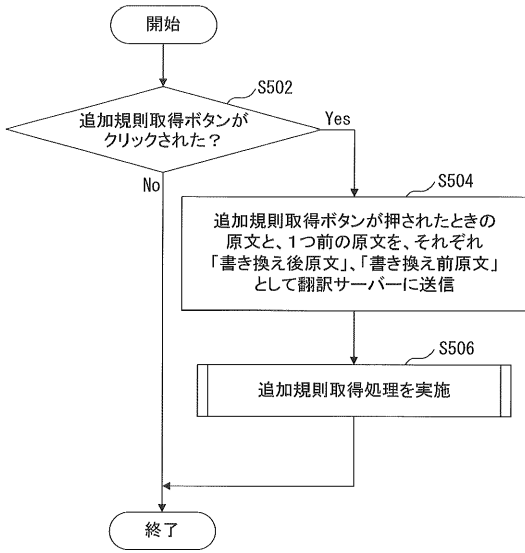
【 図 2 3 】

第2の実施形態の動作説明図



【 図 2 4 】

第2の実施形態のフローチャート



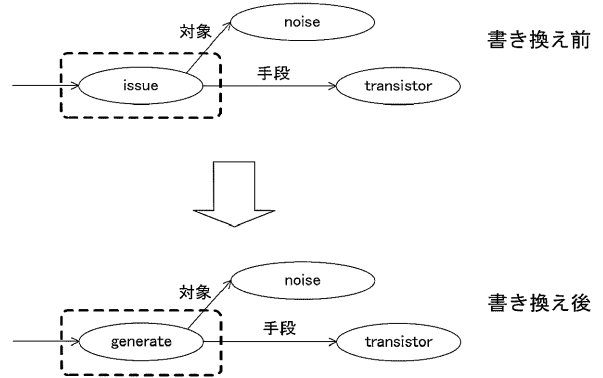
【 図 2 5 】

第3の実施形態の動作説明図

書換え前: 「The noise is **issued** by the transistors」  
 (=雑音はトランジスタによって発行される)  
 書換え後: 「The noise is **generated** by the transistors」  
 (=雑音はトランジスタによって発生する)

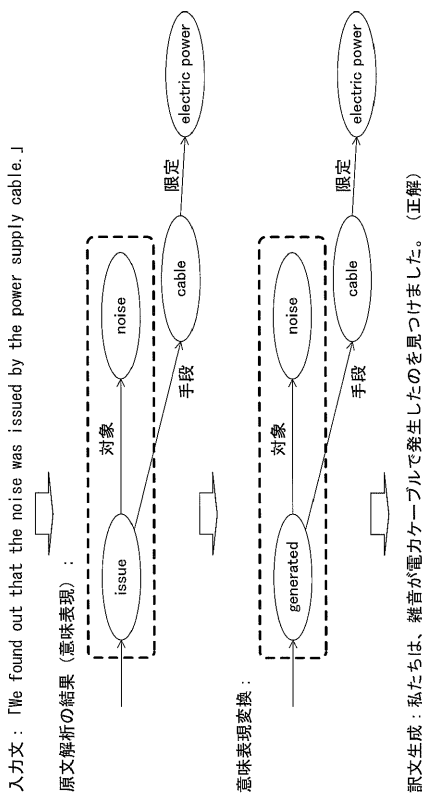
【 図 2 6 】

原文解析の結果を示す図



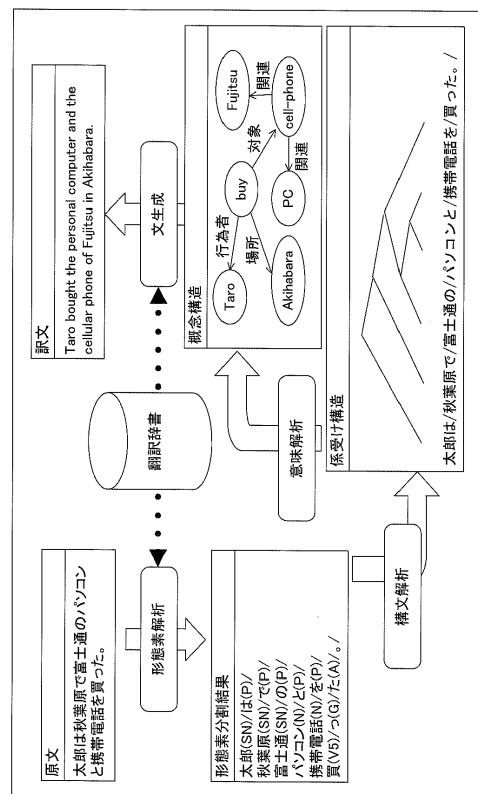
【 図 2 7 】

第3の実施形態の動作説明図



【 図 2 8 】

意味表現を説明するための図





フロントページの続き

Fターム(参考) 5B091 AA15 CA01 DA02 DA07 EA01 EA25  
5B109 QA14