

(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2009-193179

(P2009-193179A)

(43) 公開日 平成21年8月27日(2009.8.27)

(51) Int.Cl.
G06F 17/28 (2006.01)

F I
G06F 17/28

テーマコード(参考)
5B091

審査請求 未請求 請求項の数 10 O L (全 30 頁)

(21) 出願番号 特願2008-31188(P2008-31188)
(22) 出願日 平成20年2月13日(2008.2.13)

(71) 出願人 301022471
独立行政法人情報通信研究機構
東京都小金井市貫井北町4-2-1
(74) 代理人 100115749
弁理士 谷川 英和
(74) 代理人 100121223
弁理士 森本 悟道
(72) 発明者 田仲 理恵
東京都小金井市貫井北町4-2-1 独立
行政法人情報通信研究機構内
(72) 発明者 石田 亨
東京都小金井市貫井北町4-2-1 独立
行政法人情報通信研究機構内

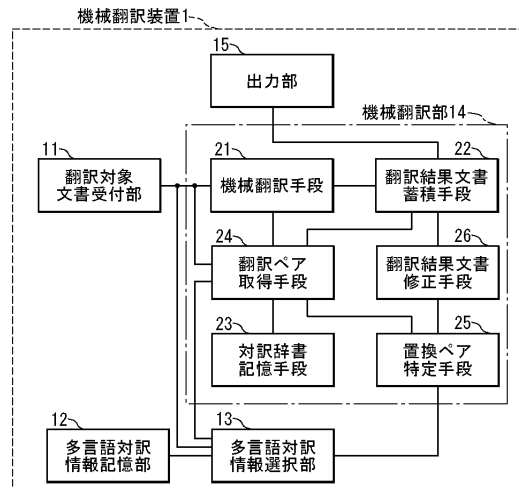
最終頁に続く

(54) 【発明の名称】 機械翻訳装置、機械翻訳方法、及びプログラム

(57) 【要約】

【課題】二言語間の機械翻訳を繰り返すことによって、第1言語から第N言語(Nは3以上の整数である)までの翻訳を行う機械翻訳装置を提供する。

【解決手段】第1言語の翻訳対象となる翻訳対象文書を受け付ける翻訳対象文書受付部11、第1言語から第N言語までの同義の単語の組である多言語対訳情報が1以上記憶される多言語対訳情報記憶部12、その1以上の多言語対訳情報から、第i言語(iは1からN-1の整数である)の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する多言語対訳情報選択部13、選択後の多言語対訳情報に含まれる二言語間の対訳関係が用いられるように、第1言語の翻訳対象文書からはじめて、第i言語の翻訳対象文書を第(i+1)言語に機械翻訳する処理を第N言語まで繰り返す機械翻訳部14、機械翻訳後の第N言語の文書を出力する出力部15を備える。



【選択図】 図1

【特許請求の範囲】

【請求項 1】

二言語間の機械翻訳を繰り返すことによって、第 1 言語から第 N 言語（N は 3 以上の整数である）までの翻訳を行う機械翻訳装置であって、

第 1 言語の翻訳対象となる文書である翻訳対象文書を受け付ける翻訳対象文書受付部と、第 1 言語から第 N 言語までの同義の単語の組である多言語対訳情報が 1 以上記憶される多言語対訳情報記憶部と、

前記多言語対訳情報記憶部で記憶されている 1 以上の多言語対訳情報から、第 i 言語（i は 1 から N - 1 の整数である）の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する多言語対訳情報選択部と、

前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる二言語間の対訳関係が利用されるように、前記翻訳対象文書受付部が受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 (i + 1) 言語に機械翻訳する処理を第 N 言語への機械翻訳を行うまで繰り返す機械翻訳部と、

前記機械翻訳部が機械翻訳した第 N 言語の文書を出力する出力部と、を備えた機械翻訳装置。

【請求項 2】

前記機械翻訳部は、

前記翻訳対象文書受付部が受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 (i + 1) 言語に機械翻訳する処理を繰り返す機械翻訳手段と、

前記機械翻訳手段が機械翻訳した結果の文書である翻訳結果文書を蓄積する翻訳結果文書蓄積手段と、

翻訳対象文書に含まれる単語と、当該翻訳対象文書を前記機械翻訳手段が機械翻訳した翻訳結果文書に含まれる単語とのペアであって、対訳関係にある単語のペアである翻訳ペアを取得する翻訳ペア取得手段と、

前記翻訳ペア取得手段が取得した翻訳ペアのうち、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれない翻訳ペアに含まれる目的言語の単語である置換対象単語と、当該翻訳ペアに含まれる原言語の単語を含む多言語対訳情報であって、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる前記目的言語の単語である置換結果単語とのペアである置換ペアを特定する置換ペア特定手段と、

前記翻訳結果文書蓄積手段が蓄積した翻訳結果文書に含まれる単語のうち、前記置換ペア特定手段が特定した置換ペアに含まれる置換対象単語を、当該置換ペアに含まれる置換結果単語に置換した文書である修正翻訳結果文書を生成する翻訳結果文書修正手段と、を備え、

前記機械翻訳手段は、前記翻訳結果文書修正手段が生成した修正翻訳結果文書を翻訳対象文書として機械翻訳し、

前記出力部は、前記翻訳結果文書修正手段が生成した第 N 言語の修正翻訳結果文書を出力する、請求項 1 記載の機械翻訳装置。

【請求項 3】

前記機械翻訳部は、

第 i 言語の単語と第 (i + 1) 言語の単語とを対応付ける情報である対訳辞書が記憶される対訳辞書記憶手段をさらに備え、

前記翻訳ペア取得手段は、前記対訳辞書記憶手段で記憶されている対訳辞書を用いて、翻訳ペアを取得する、請求項 2 記載の機械翻訳装置。

【請求項 4】

前記翻訳ペア取得手段は、前記機械翻訳手段から翻訳ペアを取得する、請求項 2 記載の機械翻訳装置。

【請求項 5】

前記翻訳ペア取得手段は、原言語の単語が、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる翻訳ペアを取得する、請求項 1 から請求項 4 のいずれか記載の機械翻

10

20

30

40

50

訳装置。

【請求項 6】

前記多言語対訳情報選択部は、二言語間の機械翻訳ごとに、前回の選択結果である多言語対訳情報からの多言語対訳情報の選択を行う、請求項 1 から請求項 5 のいずれか記載の機械翻訳装置。

【請求項 7】

翻訳対象文書受付部と、第 1 言語から第 N 言語 (N は 3 以上の整数である) までの同義の単語の組である多言語対訳情報が 1 以上記憶される多言語対訳情報記憶部と、多言語対訳情報選択部と、機械翻訳部と、出力部とを用いて、二言語間の機械翻訳を繰り返すことによって、第 1 言語から第 N 言語までの翻訳を行う機械翻訳方法であって、

前記翻訳対象文書受付部が、第 1 言語の翻訳対象となる文書である翻訳対象文書を受け付ける翻訳対象文書受付ステップと、

前記多言語対訳情報選択部が、前記多言語対訳情報記憶部で記憶されている 1 以上の多言語対訳情報から、第 i 言語 (i は 1 から N - 1 の整数である) の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する多言語対訳情報選択ステップと、

前記機械翻訳部が、前記多言語対訳情報選択ステップで選択した多言語対訳情報に含まれる二言語間の対訳関係が用いられるように、前記翻訳対象文書受付ステップで受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 (i + 1) 言語に機械翻訳する処理を第 N 言語への機械翻訳を行うまで繰り返す第 1 の機械翻訳ステップと、前記出力部が、前記第 1 の機械翻訳ステップで機械翻訳した第 N 言語の文書を出力する出力ステップと、を備えた機械翻訳方法。

【請求項 8】

前記機械翻訳部は、機械翻訳手段と、翻訳結果文書蓄積手段と、翻訳ペア取得手段と、置換ペア特定手段と、翻訳結果文書修正手段とを備え、

前記第 1 の機械翻訳ステップは、

前記機械翻訳手段が、前記翻訳対象文書受付ステップで受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 (i + 1) 言語に機械翻訳する処理を繰り返す第 2 の機械翻訳ステップと、

前記翻訳結果文書蓄積手段が、前記第 2 の機械翻訳ステップで機械翻訳した結果の文書である翻訳結果文書を蓄積する翻訳結果文書蓄積ステップと、

前記翻訳ペア取得手段が、翻訳対象文書に含まれる単語と、当該翻訳対象文書を前記第 2 の機械翻訳ステップで機械翻訳した翻訳結果文書に含まれる単語とのペアであって、対訳関係にある単語のペアである翻訳ペアを取得する翻訳ペア取得ステップと、

前記置換ペア特定手段が、前記翻訳ペア取得ステップで取得した翻訳ペアのうち、前記多言語対訳情報選択ステップで選択した多言語対訳情報に含まれない翻訳ペアに含まれる目的言語の単語である置換対象単語と、当該翻訳ペアに含まれる原言語の単語を含む多言語対訳情報であって、前記多言語対訳情報選択ステップで選択した多言語対訳情報に含まれる前記目的言語の単語である置換結果単語とのペアである置換ペアを特定する置換ペア特定ステップと、

前記翻訳結果文書修正手段が、前記翻訳結果文書蓄積ステップで蓄積した翻訳結果文書に含まれる単語のうち、前記置換ペア特定ステップで特定した置換ペアに含まれる置換対象単語を、当該置換ペアに含まれる置換結果単語に置換した文書である修正翻訳結果文書を生成する翻訳結果文書修正ステップと、を備え、

前記第 2 の機械翻訳ステップでは、前記翻訳結果文書修正ステップで生成した修正翻訳結果文書を翻訳対象文書として機械翻訳し、

前記出力ステップでは、前記翻訳結果文書修正ステップで生成した第 N 言語の修正翻訳結果文書を出力する、請求項 7 記載の機械翻訳方法。

【請求項 9】

コンピュータを、

二言語間の機械翻訳を繰り返すことによって、第 1 言語から第 N 言語 (N は 3 以上の整数

10

20

30

40

50

である)までの翻訳を行う機械翻訳装置として機能させるためのプログラムであって、第1言語の翻訳対象となる文書である翻訳対象文書を受け付ける翻訳対象文書受付部と、第1言語から第N言語までの同義の単語の組である多言語対訳情報が1以上記憶される多言語対訳情報記憶部で記憶されている1以上の多言語対訳情報から、第i言語(iは1からN-1の整数である)の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する多言語対訳情報選択部と、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる二言語間の対訳関係が利用されるように、前記翻訳対象文書受付部が受け付けた第1言語の翻訳対象文書からはじめて、第i言語の翻訳対象文書を第(i+1)言語に機械翻訳する処理を第N言語への機械翻訳を行うまで繰り返す機械翻訳部と、前記機械翻訳部が機械翻訳した第N言語の文書を出力する出力部として機能させるためのプログラム。

10

【請求項10】

前記機械翻訳部は、

前記翻訳対象文書受付部が受け付けた第1言語の翻訳対象文書からはじめて、第i言語の翻訳対象文書を第(i+1)言語に機械翻訳する処理を繰り返す機械翻訳手段と、

前記機械翻訳手段が機械翻訳した結果の文書である翻訳結果文書を蓄積する翻訳結果文書蓄積手段と、

翻訳対象文書に含まれる単語と、当該翻訳対象文書を前記機械翻訳手段が機械翻訳した翻訳結果文書に含まれる単語とのペアであって、対訳関係にある単語のペアである翻訳ペアを取得する翻訳ペア取得手段と、

20

前記翻訳ペア取得手段が取得した翻訳ペアのうち、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれない翻訳ペアに含まれる目的言語の単語である置換対象単語と、当該翻訳ペアに含まれる原言語の単語を含む、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる前記目的言語の単語である置換結果単語とのペアである置換ペアを特定する置換ペア特定手段と、

前記翻訳結果文書蓄積手段が蓄積した翻訳結果文書に含まれる単語のうち、前記置換ペア特定手段が特定した置換ペアに含まれる置換対象単語を、当該置換ペアに含まれる置換結果単語に置換した文書である修正翻訳結果文書を生成する翻訳結果文書修正手段と、を備え、

30

前記機械翻訳手段は、前記翻訳結果文書修正手段が生成した修正翻訳結果文書を翻訳対象文書として機械翻訳し、

前記出力部は、前記翻訳結果文書修正手段が生成した第N言語の修正翻訳結果文書を出力する、請求項9記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、二言語間の機械翻訳を繰り返すことにより、第1言語から第N言語(Nは3以上の整数)までの機械翻訳を行う機械翻訳装置等に関する。

【背景技術】

40

【0002】

従来、原言語の文書を目的言語の文書に機械的に翻訳する機械翻訳装置が知られており、その機械翻訳の精度が向上してきている(例えば、特許文献1参照)。

【0003】

そのような二言語間の機械翻訳を繰り返すことによって、機械翻訳を行うことができなかった二言語間の機械翻訳を実現することも可能である。例えば、日英機械翻訳と、英独機械翻訳は存在するが、日独機械翻訳が存在しない場合であっても、日本語の文書に対して日英機械翻訳を実行して英文の翻訳文を取得し、その英文の翻訳文に対して英独機械翻訳を実行してドイツ語の翻訳文を取得することによって、日本語からドイツ語への機械翻訳を実現することも可能である。

50

【特許文献1】特開2008-15844

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかしながら、そのような二言語間の機械翻訳を繰り返す場合には、単語の多義性により、訳語がドリフトしてしまう可能性がある。例えば、日本語の単語「過ち」が、日英機械翻訳によって英語の単語「fault」に機械翻訳され、その英語の単語が英独機械翻訳によってドイツ語の単語「Schuld」に機械翻訳されることがある。このドイツ語の単語「Schuld」は、「責任」という意味である。したがって、この機械翻訳では、日本語の単語「過ち」が異なる意味に翻訳されており、訳語にドリフトが発生している。これは、英語の単語「fault」には、「過ち」という意味と、「責任」という意味があるために起こることである。

10

【0005】

このように、二言語間の機械翻訳を繰り返すことによって第1言語から第N言語（Nは3以上の整数）までの翻訳を行う機械翻訳を行う場合には、単語の多義性に起因する訳語のドリフトが発生することがあり、その結果として、第1言語の文書の意味と、翻訳後の第N言語の文書の意味とが異なるものになってしまうことがありうる。

【0006】

本発明は、このような問題を解決するためになされたものであり、二言語間の機械翻訳を繰り返すことによって第1言語から第N言語（Nは3以上の整数）までの翻訳を行う機械翻訳を行う場合であっても、訳語のドリフトの発生を抑制することができる機械翻訳装置等を提供することを目的とする。

20

【課題を解決するための手段】

【0007】

上記目的を達成するため、本発明による機械翻訳装置は、二言語間の機械翻訳を繰り返すことによって、第1言語から第N言語（Nは3以上の整数である）までの翻訳を行う機械翻訳装置であって、第1言語の翻訳対象となる文書である翻訳対象文書を受け付ける翻訳対象文書受付部と、第1言語から第N言語までの同義の単語の組である多言語対訳情報が1以上記憶される多言語対訳情報記憶部と、前記多言語対訳情報記憶部で記憶されている1以上の多言語対訳情報から、第i言語（iは1からN-1の整数である）の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する多言語対訳情報選択部と、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる二言語間の対訳関係が用いられるように、前記翻訳対象文書受付部が受け付けた第1言語の翻訳対象文書からはじめて、第i言語の翻訳対象文書を第(i+1)言語に機械翻訳する処理を第N言語への機械翻訳を行うまで繰り返す機械翻訳部と、前記機械翻訳部が機械翻訳した第N言語の文書を出力する出力部と、を備えたものである。

30

【0008】

このような構成により、二言語間の機械翻訳を繰り返すことによって、第1言語から第N言語までの翻訳を行うことができる。その機械翻訳の際に、多言語対訳情報選択部によって選択された多言語対訳情報を用いることによって、訳語のドリフトの発生を抑制することができる。

40

【0009】

また、本発明による機械翻訳装置では、前記機械翻訳部は、前記翻訳対象文書受付部が受け付けた第1言語の翻訳対象文書からはじめて、第i言語の翻訳対象文書を第(i+1)言語に機械翻訳する処理を繰り返す機械翻訳手段と、前記機械翻訳手段が機械翻訳した結果の文書である翻訳結果文書を蓄積する翻訳結果文書蓄積手段と、翻訳対象文書に含まれる単語と、当該翻訳対象文書を前記機械翻訳手段が機械翻訳した翻訳結果文書に含まれる単語とのペアであって、対訳関係にある単語のペアである翻訳ペアを取得する翻訳ペア取得手段と、前記翻訳ペア取得手段が取得した翻訳ペアのうち、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれない翻訳ペアに含まれる目的言語の単語である置換

50

対象単語と、当該翻訳ペアに含まれる原言語の単語を含む多言語対訳情報であって、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる前記目的言語の単語である置換結果単語とのペアである置換ペアを特定する置換ペア特定手段と、前記翻訳結果文書蓄積手段が蓄積した翻訳結果文書に含まれる単語のうち、前記置換ペア特定手段が特定した置換ペアに含まれる置換対象単語を、当該置換ペアに含まれる置換結果単語に置換した文書である修正翻訳結果文書を生成する翻訳結果文書修正手段と、を備え、前記機械翻訳手段は、前記翻訳結果文書修正手段が生成した修正翻訳結果文書を翻訳対象文書として機械翻訳し、前記出力部は、前記翻訳結果文書修正手段が生成した第N言語の修正翻訳結果文書を出力してもよい。

【0010】

このような構成により、汎用の機械翻訳手段を用いた機械翻訳において、翻訳結果文書に含まれる単語が選択後の多言語対訳情報に含まれる単語となるように修正することができ、訳語のドリフトの発生を抑制することができる。

【0011】

また、本発明による機械翻訳装置では、前記機械翻訳部は、第i言語の単語と第(i+1)言語の単語とを対応付ける情報である対訳辞書が記憶される対訳辞書記憶手段をさらに備え、前記翻訳ペア取得手段は、前記対訳辞書記憶手段で記憶されている対訳辞書を用いて、翻訳ペアを取得してもよい。

【0012】

このような構成により、翻訳ペア取得手段は、例えば、機械翻訳手段から翻訳ペアを受け取ることができない場合であっても、翻訳対象文書と、翻訳結果文書と、対訳辞書とを用いて、翻訳ペアを取得することができる。

【0013】

また、本発明による機械翻訳装置では、前記翻訳ペア取得手段は、前記機械翻訳手段から翻訳ペアを取得してもよい。

【0014】

また、本発明による機械翻訳装置では、前記翻訳ペア取得手段は、原言語の単語が、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる翻訳ペアを取得してもよい。

【0015】

このような構成により、翻訳ペア取得手段は、必要十分な量の翻訳ペアを取得することとなり、余分な翻訳ペアの取得を回避することができる。その結果、翻訳ペアを保持しておく記録領域を削減することができる。

【0016】

また、本発明による機械翻訳装置では、前記多言語対訳情報選択部は、二言語間の機械翻訳ごとに、前回の選択結果である多言語対訳情報からの多言語対訳情報の選択を行ってもよい。

【0017】

このような構成により、二言語間の機械翻訳が行われるごとに、多言語対訳情報を絞り込んでいくことができ、後段の二言語間の機械翻訳になるほど、より翻訳の処理が速くなりうる。

【発明の効果】

【0018】

本発明による機械翻訳装置等によれば、二言語間の機械翻訳を繰り返すことによって第1言語から第N言語(Nは3以上の整数)までの翻訳を行う機械翻訳を行う場合であっても、訳語のドリフトの発生を抑制することができる。

【発明を実施するための最良の形態】

【0019】

以下、本発明による機械翻訳装置について、実施の形態を用いて説明する。なお、以下の実施の形態において、同じ符号を付した構成要素及びステップは同一または相当するも

10

20

30

40

50

のであり、再度の説明を省略することがある。

【0020】

(実施の形態1)

本発明の実施の形態1による機械翻訳装置について、図面を参照しながら説明する。本実施の形態による機械翻訳装置は、二言語間の機械翻訳を繰り返すことによって、第1言語から第N言語(Nは3以上の整数である)までの機械翻訳を行うものである。

【0021】

図1は、本実施の形態による機械翻訳装置1の構成を示すブロック図である。本実施の形態による機械翻訳装置1は、翻訳対象文書受付部11と、多言語対訳情報記憶部12と、多言語対訳情報選択部13と、機械翻訳部14と、出力部15とを備える。

10

【0022】

翻訳対象文書受付部11は、第1言語の翻訳対象となる文書である翻訳対象文書を受け付ける。翻訳対象文書は、例えば、一文であってもよく、複数の文を含むものであってもよく、文の一部(例えば、フレーズ等)であってもよい。翻訳対象文書は、翻訳対象を特定することができるデータであればどのようなデータであってもよく、例えば、テキストデータであってもよい。

【0023】

翻訳対象文書受付部11は、例えば、入力デバイス(例えば、キーボードやマウス、タッチパネルなど)から入力された翻訳対象文書を受け付けてもよく、有線もしくは無線の通信回線を介して送信された翻訳対象文書を受信してもよく、所定の記録媒体(例えば、光ディスクや磁気ディスク、半導体メモリなど)から読み出された翻訳対象文書を受け付けてもよい。なお、翻訳対象文書受付部11は、受け付けを行うためのデバイス(例えば、モデムやネットワークカードなど)を含んでもよく、あるいは含まなくてもよい。また、翻訳対象文書受付部11は、ハードウェアによって実現されてもよく、あるいは所定のデバイスを駆動するドライバ等のソフトウェアによって実現されてもよい。

20

【0024】

多言語対訳情報記憶部12では、第1言語から第N言語までの同義の単語の組である多言語対訳情報が1以上記憶される。このように、多言語対訳情報には、同義の単語である、第1言語の単語と、第2言語の単語と、...、第(N-1)言語の単語と、第N言語の単語とが含まれる。多言語対訳情報は、例えば、日本語の単語「空」と、英語の単語「sky」と、ドイツ語の単語「Himmel」とを含む情報であってもよい。ここで、単語は、文を構成する単位であり、例えば、文法上の意味での単語であってもよく、形態素であってもよく、文法上の意味での単語の少数の並び(イディオム)であってもよい。この多言語対訳情報が生成される方法は問わない。例えば、多言語対訳情報は、人手によって生成されてもよく、あるいは、機械的に生成されてもよい。なお、多言語対訳情報記憶部12で記憶されている多言語対訳情報の個数は、2以上であることが好適である。より多い多言語対訳情報が記憶されていた方が、後述する多言語対訳情報選択部13による選択の幅がより広がることになり、好適だからである。本実施の形態では、多言語対訳情報記憶部12で2以上の多言語対訳情報が記憶されている場合について主に説明する。

30

【0025】

多言語対訳情報記憶部12に一または複数の多言語対訳情報が記憶される過程は問わない。例えば、記録媒体を介して一または複数の多言語対訳情報が多言語対訳情報記憶部12で記憶されるようになってもよく、通信回線等を介して送信された一または複数の多言語対訳情報が多言語対訳情報記憶部12で記憶されるようになってもよく、あるいは、入力デバイスを介して入力された一または複数の多言語対訳情報が多言語対訳情報記憶部12で記憶されるようになってもよい。多言語対訳情報記憶部12での記憶は、RAM等における一時的な記憶でもよく、あるいは、長期的な記憶でもよい。多言語対訳情報記憶部12は、所定の記録媒体(例えば、半導体メモリや磁気ディスク、光ディスクなど)によって実現されうる。

40

【0026】

50

多言語対訳情報選択部 1 3 は、多言語対訳情報記憶部 1 2 で記憶されている 1 以上の多言語対訳情報から、第 i 言語 (i は 1 から $N - 1$ の整数である) の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する。一の多言語対訳情報からの選択の際には、その多言語対訳情報を採用するかどうかの処理となる。また、前述のように、より多くの多言語対訳情報から選択を行う方が好適である。また、多言語対訳情報選択部 1 3 は、二言語間の機械翻訳ごとに、前回の選択結果である多言語対訳情報からの多言語対訳情報の選択を行う。したがって、第 i 言語の翻訳対象文書が第 ($i + 1$) 言語の文書に機械翻訳されるごとに、この多言語対訳情報選択部 1 3 による選択が行われることになる。前述のように、多言語対訳情報は、第 1 言語から第 N 言語までの単語の組であり、多言語対訳情報選択部 1 3 は、その多言語対訳情報に含まれる第 i 言語の単語が、第 i 言語の翻訳対象文書に含まれる多言語対訳情報を選択することになる。したがって、二言語間の機械翻訳が繰り返して実行されるごとに、選択後の多言語対訳情報が絞り込まれていくことになる。

10

【 0 0 2 7 】

多言語対訳情報選択部 1 3 は、選択後の多言語対訳情報を図示しない記録媒体で一時的に記憶してもよく、あるいは、多言語対訳情報記憶部 1 2 で記憶されている多言語対訳情報のうち、選択されたものについてフラグ等を付加することによって、選択後の多言語対訳情報を特定できるようにしてもよい。このように、選択後の多言語対訳情報を特定することができるのであれば、その選択後の多言語対訳情報を示す方法は問わない。

【 0 0 2 8 】

機械翻訳部 1 4 は、多言語対訳情報選択部 1 3 が選択した多言語対訳情報に含まれる二言語間の対訳関係が用いられるように、翻訳対象文書受付部 1 1 が受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 ($i + 1$) 言語に機械翻訳する処理を第 N 言語への機械翻訳を行うまで繰り返す。「多言語対訳情報選択部 1 3 が選択した多言語対訳情報に含まれる二言語間の対訳関係が用いられるように機械翻訳する」とは、例えば、機械翻訳部 1 4 が第 i 言語から第 ($i + 1$) 言語への機械翻訳を行う場合に、第 i 言語の翻訳対象文書に含まれる単語が選択後の多言語対訳情報に含まれているのであれば、その単語が、その多言語対訳情報に含まれている第 ($i + 1$) 言語の単語に翻訳されるように機械翻訳することである。

20

【 0 0 2 9 】

機械翻訳部 1 4 は、選択後の多言語対訳情報を取り込み、機械翻訳の仕組み自体を変更することによって、そのような機械翻訳がなされるようにしてもよく、あるいは、機械翻訳自体は従来のもを用い、その機械翻訳されたものを、選択後の多言語対訳情報を用いて修正するようにしてもよい。本実施の形態では、後者の場合について説明する。後者の場合には、機械翻訳部 1 4 は、図 1 で示されるように、機械翻訳手段 2 1 と、翻訳結果文書蓄積手段 2 2 と、対訳辞書記憶手段 2 3 と、翻訳ペア取得手段 2 4 と、置換ペア特定手段 2 5 と、翻訳結果文書修正手段 2 6 とを備える。

30

【 0 0 3 0 】

機械翻訳手段 2 1 は、翻訳対象文書受付部 1 1 が受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 ($i + 1$) 言語に機械翻訳する処理を繰り返して実行する。この機械翻訳手段 2 1 は、 i が 1 から ($N - 1$) のすべての場合について、第 i 言語から第 ($i + 1$) 言語への機械翻訳を行うものであり、その機械翻訳の仕組み自体は、従来 of 機械翻訳を用いるものである。したがって、機械翻訳手段 2 1 は、選択後の多言語対訳情報を考慮しないで機械翻訳を行うことになる。例えば、機械翻訳部 1 4 が、日本語から英語への機械翻訳と、英語からドイツ語への機械翻訳とを行う場合には、機械翻訳手段 2 1 は、日英機械翻訳と、英独機械翻訳とを行う。なお、機械翻訳手段 2 1 は、後述する翻訳結果文書修正手段 2 6 が生成した修正翻訳結果文書を翻訳対象文書として機械翻訳する。また、機械翻訳手段 2 1 が機械翻訳を行う対象となる文書を翻訳対象文書と呼び、機械翻訳手段 2 1 が機械翻訳を行った結果の文書を翻訳結果文書と呼ぶ。

40

【 0 0 3 1 】

翻訳結果文書蓄積手段 2 2 は、機械翻訳手段 2 1 が機械翻訳した結果の文書である翻訳

50

結果文書を図示しない記録媒体に蓄積する。この記録媒体は、例えば、半導体メモリや、光ディスク、磁気ディスク等であり、翻訳結果文書蓄積手段 2 2 が有していてもよく、あるいは翻訳結果文書蓄積手段 2 2 の外部に存在してもよい。また、この記録媒体は、翻訳結果文書を一時的に記憶するものであってもよく、そうでなくてもよい。

【 0 0 3 2 】

対訳辞書記憶手段 2 3 では、 i が 1 から $(N - 1)$ のすべての場合について、第 i 言語の単語と第 $(i + 1)$ 言語の単語とを対応付ける情報である対訳辞書が記憶される。この対訳辞書では、対訳関係にある二言語の単語を対応付ける情報である。

【 0 0 3 3 】

ここで、「二言語の単語を対応付ける」とは、ある言語の単語、他の言語の単語の一方の単語から、他方の単語を取得できればよいという意味である。したがって、対訳辞書は、ある言語の単語と他の言語の単語とを組として含む情報を有してもよく、ある言語の単語と他の言語の単語とをリンク付ける情報であってもよい。後者の場合には、対訳辞書は、例えば、ある言語の単語と他の言語の単語の格納されている位置を示すポインタやアドレスとを対応付ける情報であってもよい。本実施の形態では、前者の場合について説明する。

10

【 0 0 3 4 】

なお、対訳辞書では、対訳関係にある単語について、原言語の単語に対して、目的言語の単語が 1 以上対応付けられていることが好適である。すなわち、対訳辞書は、例えば、原言語である日本語の単語「空(そら)」と、目的言語である英語の単語「s k y , a i r , h e a v e n」とを含む組を有してもよい。

20

【 0 0 3 5 】

例えば、機械翻訳部 1 4 が、日本語から英語への機械翻訳と、英語からドイツ語への機械翻訳とを行う場合には、対訳辞書記憶手段 2 3 では、日英対訳辞書と、英独対訳辞書とが記憶されることになる。このように、対訳辞書記憶手段 2 3 では、複数の種類の対訳辞書が記憶されることになる。

【 0 0 3 6 】

対訳辞書記憶手段 2 3 に複数の種類の対訳辞書が記憶される過程は問わない。例えば、記録媒体を介して複数の種類の対訳辞書が対訳辞書記憶手段 2 3 で記憶されるようになってよく、通信回線等を介して送信された複数の種類の対訳辞書が対訳辞書記憶手段 2 3 で記憶されるようになってよく、あるいは、入力デバイスを介して入力された複数の種類の対訳辞書が対訳辞書記憶手段 2 3 で記憶されるようになってよく、あるいは、長期的な記憶でもよい。対訳辞書記憶手段 2 3 は、所定の記録媒体(例えば、半導体メモリや磁気ディスク、光ディスクなど)によって実現されうる。

30

【 0 0 3 7 】

翻訳ペア取得手段 2 4 は、翻訳ペアを取得する。ここで、翻訳ペアとは、翻訳対象文書に含まれる単語と、その翻訳対象文書を機械翻訳手段 2 1 が機械翻訳した翻訳結果文書に含まれる単語とのペアである。その翻訳ペアに含まれるペアとなる単語は、対訳関係にある単語のペアである。翻訳ペア取得手段 2 4 は、機械翻訳手段 2 1 から翻訳ペアを取得してもよく、あるいは、対訳辞書記憶手段 2 3 で記憶されている対訳辞書を用いて、翻訳ペアを取得してもよい。機械翻訳手段 2 1 は、通常、機械翻訳の際に、原言語の単語と、その単語の翻訳後の単語である目的言語の単語とを特定することができる。したがって、前者の場合には、翻訳ペア取得手段 2 4 は、その原言語の単語と目的言語の単語とのペアである翻訳ペアを取得するものである。一方、翻訳ペアを機械翻訳手段 2 1 から取得することができない場合には、後者のように、翻訳ペア取得手段 2 4 は、対訳辞書を用いて翻訳ペアを取得する。本実施の形態では、後者の場合、すなわち、翻訳ペア取得手段 2 4 が対訳辞書を用いて翻訳ペアを取得する場合について説明する。

40

【 0 0 3 8 】

具体的には、翻訳ペア取得手段 2 4 は、第 i 言語の翻訳対象文書に含まれる単語を、第

50

i 言語と第 (i + 1) 言語間の対訳辞書によって第 (i + 1) 言語の単語に翻訳した単語が、第 (i + 1) 言語の翻訳結果文書に含まれる場合に、第 i 言語の翻訳対象文書に含まれるその単語と、その単語を対訳辞書によって第 (i + 1) 言語の単語に翻訳した単語とのペアである翻訳ペアを取得する。

【 0 0 3 9 】

また、翻訳ペア取得手段 2 4 は、翻訳ペアに含まれる原言語の単語が、多言語対訳情報選択部 1 3 が選択した多言語対訳情報に含まれる翻訳ペアを取得してもよい。なお、翻訳ペアは、第 i 言語の単語と、第 (i + 1) 言語の単語とのペアであるが、そのうち、第 i 言語を原言語と呼び、第 (i + 1) 言語を目的言語と呼ぶことにする。翻訳ペア取得手段 2 4 が、翻訳ペアに含まれる原言語の単語が選択後の多言語対訳情報に含まれる翻訳ペアを取得する場合には、例えば、翻訳ペア取得手段 2 4 は、前述のようにして翻訳ペアを取得した後に、その取得した各翻訳ペアに含まれる原言語の単語が選択後の多言語対訳情報に含まれるかどうか判断し、原言語の単語が選択後の多言語対訳情報に含まれる翻訳ペアを残し、原言語の単語が選択後の多言語対訳情報に含まれない翻訳ペアを破棄する (翻訳ペアでないとする) ようにしてもよい。

10

【 0 0 4 0 】

また、翻訳ペア取得手段 2 4 は、特定の品詞の単語のペアである翻訳ペアを取得してもよい。例えば、翻訳ペア取得手段 2 4 は、多言語対訳情報に名詞の単語が含まれる場合に、名詞の単語のペアである翻訳ペアを取得してもよく、多言語対訳情報に自立語の単語が含まれる場合に、自立語の単語のペアである翻訳ペアを取得してもよく、多言語対訳情報に含まれる単語の品詞と一致する品詞の単語のペアである翻訳ペアを取得してもよい。多言語対訳情報に含まれない品詞の単語のペアである翻訳ペアを取得したとしても、その翻訳ペアは後の処理で用いられないからである。なお、特定の品詞の翻訳ペアのみを取得する場合には、例えば、翻訳ペア取得手段 2 4 は、翻訳対象文書を形態素解析等して各単語の品詞を解析し、その解析結果を用いて、特定の品詞の単語についてのみ、翻訳ペアを取得する処理を行ってもよい。なお、形態素解析等の各単語の品詞を解析する手法については公知であり、詳細な説明を省略する。

20

【 0 0 4 1 】

なお、形態素解析のシステムとして、日本語の場合には、例えば、奈良先端科学技術大学院大学で開発された「Chasen (茶筌)」 (<http://chasen.naist.jp>) 等が知られている。また、英語の場合には、英単語に品詞を付与するソフトウェアとして、例えば、「TnT」 (<http://www.coli.uni-saarland.de/~thorsten/tnt/>) や「Brill Tagger」 (<http://www.cs.jhu.edu/~brill/>) 等が知られている。Brill のものについては、例えば、次の文献を参照されたい。

30

【 0 0 4 2 】

文献：Eric Brill、「Transformation - Based Error - Driven Learning and Natural Language Processing: A Case Study in Part - of - Speech Tagging」、Computational Linguistics, Vol . 21, No . 4, p . 543 - 565、1995年

40

【 0 0 4 3 】

また、翻訳ペア取得手段 2 4 は、対訳辞書を用いた翻訳ペアの取得において、翻訳対象文書に含まれる原言語の単語と対訳関係にある翻訳結果文書の単語が見つからない場合には、その原言語の単語を機械翻訳手段 2 1 によって機械翻訳した結果を用いて、その原言語の単語と対訳関係にある翻訳結果文書の単語を特定してもよい。具体的には、翻訳ペア取得手段 2 4 は、ある原言語の単語を機械翻訳手段 2 1 によって機械翻訳した結果である目的言語の単語が翻訳結果文書に含まれるのであれば、その原言語の単語と、その単語の機械翻訳結果である目的言語の単語とをペアとして含む翻訳ペアを取得してもよい。

【 0 0 4 4 】

50

翻訳ペア取得手段 2 4 は、取得した翻訳ペアを図示しない記録媒体で一時的に記憶してもよく、あるいは、対訳辞書記憶手段 2 3 で記憶されている対訳辞書に含まれる情報において、取得した翻訳ペアに該当する単語にフラグ等を付加することによって、翻訳ペアを特定できるようにしてもよい。このように、翻訳ペアを特定することができるのであれば、その翻訳ペアを示す方法は問わない。

【 0 0 4 5 】

置換ペア特定手段 2 5 は、置換ペアを特定する。この置換ペアは、置換対象単語と、置換結果単語とのペアである。置換対象単語は、翻訳ペア取得手段 2 4 が取得した翻訳ペアのうち、多言語対訳情報選択部 1 3 が選択した多言語対訳情報に含まれない翻訳ペアに含まれる目的言語の単語である。「多言語対訳情報選択部 1 3 が選択した多言語対訳情報に含まれない翻訳ペア」とは、その翻訳ペアに含まれる原言語の単語と、目的言語の単語の両方を含む選択後の多言語対訳情報が存在しない翻訳ペアである。置換結果単語は、選択後のいずれの多言語対訳情報にも含まれない翻訳ペアに含まれる原言語の単語を含む選択後の多言語対訳情報に含まれる目的言語（選択後のいずれの多言語対訳情報にも含まれない翻訳ペアの目的言語である）の単語である。なお、選択後のいずれの多言語対訳情報にも含まれない一翻訳ペアが存在した場合に、置換ペアを特定できる場合と、できない場合とがある。後者の場合としては、選択後のいずれの多言語対訳情報にも含まれない翻訳ペアに含まれる原言語の単語を含む選択後の多言語対訳情報が存在しなかった場合である。なお、翻訳ペアの取得の際に、翻訳ペアに含まれる原言語の単語が選択後の多言語対訳情報に含まれる翻訳ペアのみを取得している場合には、置換ペアを特定できない場合はないことになる。また、置換ペアを特定できる場合でも、一置換ペアのみを特定できる場合と、2 以上の置換ペアを特定できる場合とがある。後者の場合には、2 以上の置換ペアを特定してもよく、あるいは、その 2 以上の置換ペアから選択された一置換ペアのみを特定してもよい。本実施の形態では、置換ペア特定手段 2 5 が、一置換ペアのみを特定する場合について説明する。

10

20

【 0 0 4 6 】

ここで、置換ペア特定手段 2 5 が、2 以上の置換ペアから選択された一置換ペアのみを特定する方法について説明する。

【 0 0 4 7 】

[単語の出現頻度を用いる方法]

置換ペア特定手段 2 5 は、同一の置換対象単語に対応する複数の置換結果単語が存在する場合に、その置換結果単語のうち、最も出現頻度の高い置換結果単語を含む置換ペアを特定してもよい。置換ペア特定手段 2 5 は、例えば、あらかじめ図示しない記録媒体で記憶されている、単語と出現頻度を示す情報とを対応付けて有する情報を用いて、単語の出現頻度を取得することができる。この単語の出現頻度は、例えば、所定のコーパスを用いて算出されたものであってもよく、それまでに機械翻訳されている第 ($i + 1$) 言語の文書を用いて算出されたものであってもよい。

30

【 0 0 4 8 】

[文脈情報の優先度を用いる方法]

置換ペア特定手段 2 5 は、同一の置換対象単語に対応する複数の置換結果単語が存在する場合に、翻訳対象文書の前の文または文章全体の文脈情報を用いて、その文脈により近い置換ペアを特定してもよい。例えば、あらかじめ図示しない記録媒体で記憶されている、過去の同一利用者から入力された翻訳対象文書と利用した置換ペアとを対応付けて有する情報を用いて、これまで置換対象単語がどの置換結果単語で置き換えられていたかという文脈を取得することができる。また、あらかじめ図示しない記録媒体で記憶されている、文書全体のテーマを用いて、文書全体の文脈を取得することもできる。この場合には、例えば、単語とテーマとを対応付けて有する情報を用いて、文書のテーマに対応する単語を選択し、その選択した単語を含む置換ペアを特定することができる。例えば、文書のテーマが「経済」である場合には、置換結果単語の候補としての単語「土手」「銀行」のうち、「銀行」を置換結果単語としてもよい。この場合には、例えば、単語「土手」には

40

50

、テーマ「自然」が対応しており、単語「銀行」には、テーマ「経済」が対応しているものとする。

【0049】

[その他の方法]

置換ペア特定手段25は、同一の置換対象単語に対応しうる複数の置換結果単語が存在する場合に、その複数の置換結果単語から、ランダムに選択した置換結果単語を含む置換ペアを特定してもよい。また、置換ペア特定手段25は、同一の置換対象単語に対応しうる複数の置換結果単語が存在する場合に、その複数の置換結果単語から、あらかじめ決められている単語、例えば、1番目の置換結果単語を含む置換ペアを特定してもよい。

【0050】

なお、これら以外の方法によって置換ペアを特定してもよく、置換ペアを特定する方法は、これらに限定されない。

【0051】

置換ペア特定手段25は、特定した置換ペアを図示しない記録媒体で一時的に記憶してもよく、あるいは、他の情報において、特定した置換ペアに該当する単語にフラグ等を加えることによって、置換ペアを特定できるようにしてもよい。このように、置換ペアを特定することができるのであれば、その置換ペアを示す方法は問わない。

【0052】

翻訳結果文書修正手段26は、翻訳結果文書蓄積手段22が蓄積した翻訳結果文書から、修正翻訳結果文書を生成する。すなわち、翻訳結果文書修正手段26は、翻訳結果文書に含まれる単語のうち、置換ペア特定手段25が特定した置換ペアに含まれる置換対象単語を、その置換ペアに含まれる置換結果単語に置換した文書である修正翻訳結果文書を生成する。同一の置換対象単語を有する2以上の置換ペアが存在する場合には、翻訳結果文書修正手段26は、その2以上の置換ペアからいずれかを選択し、その選択した置換ペアを用いて修正翻訳結果文書の生成を行ってもよい。

【0053】

なお、修正翻訳結果文書は、翻訳結果文書蓄積手段22が翻訳結果文書を蓄積した記録媒体において記憶されてもよく、あるいは、他の記録媒体において記憶されてもよい。また、前述のように、修正翻訳結果文書が、機械翻訳手段21によって翻訳対象文書として用いられることになる。

【0054】

なお、翻訳結果文書に、置換ペアに含まれる置換対象単語が含まれない場合、すなわち、翻訳結果情報に修正すべきところが存在しない場合には、翻訳結果文書がそのまま、翻訳結果文書修正手段26が生成した修正翻訳結果文書となるものとする。

【0055】

出力部15は、機械翻訳部14が機械翻訳した第N言語の文書を出力する。より具体的には、出力部15は、翻訳結果文書修正手段26が生成した第N言語の修正翻訳結果文書を出力する。

【0056】

ここで、この出力は、例えば、表示デバイス(例えば、CRTや液晶ディスプレイなど)への表示でもよく、所定の機器への通信回線を介した送信でもよく、プリンタによる印刷でもよく、スピーカによる音声出力でもよく、記録媒体への蓄積でもよく、他の構成要素への引き渡しでもよい。なお、出力部15は、出力を行うデバイス(例えば、表示デバイスやプリンタなど)を含んでもよく、あるいは含まなくてもよい。また、出力部15は、ハードウェアによって実現されてもよく、あるいは、それらのデバイスを駆動するドライバ等のソフトウェアによって実現されてもよい。

【0057】

なお、多言語対訳情報記憶部12と、翻訳結果文書蓄積手段22が翻訳結果文書を蓄積する記録媒体と、対訳辞書記憶手段23と、その他の各種の情報記憶される記録媒体とのうち、任意の2以上の記憶部や記録媒体は、同一の記録媒体によって実現されてもよく

10

20

30

40

50

、あるいは、別々の記録媒体によって実現されてもよい。後者の場合には、例えば、多言語対訳情報の記憶されている領域が多言語対訳情報記憶部 1 2 となり、対訳辞書の記憶されている領域が対訳辞書記憶手段 2 3 となる。

【 0 0 5 8 】

次に、本実施の形態による機械翻訳装置 1 の動作について、図 2 のフローチャートを用いて説明する。

(ステップ S 1 0 1) 翻訳対象文書受付部 1 1 は、第 1 言語の翻訳対象文書を受け付けたかどうか判断する。そして、受け付けた場合には、ステップ S 1 0 2 に進み、そうでない場合には、第 1 言語の翻訳対象文書を受け付けるまで、ステップ S 1 0 1 の処理を繰り返す。

【 0 0 5 9 】

(ステップ S 1 0 2) 機械翻訳部 1 4 は、カウンタ i を 1 に設定する。

【 0 0 6 0 】

(ステップ S 1 0 3) 多言語対訳情報選択部 1 3 は、機械翻訳部 1 4 が機械翻訳を行う第 i 言語の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する。この選択の処理以前に、すでに多言語対訳情報の選択が行われている場合には、その選択後の多言語対訳情報からの選択を行うものとする。

【 0 0 6 1 】

具体的には、多言語対訳情報選択部 1 3 は、第 i 言語の翻訳対象文書の各単語を用いて多言語対訳情報を検索し、その検索でヒットした多言語対訳情報を選択するようにしてもよい。あるいは、多言語対訳情報選択部 1 3 は、各多言語対訳情報に含まれる第 i 言語の単語で第 i 言語の翻訳対象文書を検索し、その検索でヒットした第 i 言語の単語を含む多言語対訳情報を選択するようにしてもよい。

【 0 0 6 2 】

(ステップ S 1 0 4) 機械翻訳部 1 4 は、ステップ S 1 0 3 において多言語対訳情報選択部 1 3 によって選択された多言語対訳情報を用いた第 i 言語の翻訳対象文書から、第 $(i + 1)$ 言語の文書への機械翻訳を行う。この処理の詳細については、図 3 のフローチャートを用いて後述する。

【 0 0 6 3 】

(ステップ S 1 0 5) 機械翻訳部 1 4 は、カウンタ i を 1 だけインクリメントする。

【 0 0 6 4 】

(ステップ S 1 0 6) 機械翻訳部 1 4 は、カウンタ i が N と等しいかどうか判断する。ここで、 N は、あらかじめ設定されている 3 以上の整数であり、図示しない記録媒体で記憶されているものとする。カウンタ i が N と等しい場合には、翻訳対象文書が、第 N 言語にまで翻訳されていることになる。そして、カウンタ i が N に等しい場合には、ステップ S 1 0 7 に進み、そうでない場合には、ステップ S 1 0 3 に戻る。

【 0 0 6 5 】

(ステップ S 1 0 7) 出力部 1 5 は、第 N 言語の翻訳後の文書を出力する。そして、ステップ S 1 0 1 に戻る。

【 0 0 6 6 】

なお、図 2 のフローチャートにおいて、ステップ S 1 0 6 において、カウンタ i が N に等しいかどうかを判断するのではなくてもよい。この判断は、機械翻訳部 1 4 による一連の機械翻訳を終了するかどうか判断するものであればどのようなものであってもよく、例えば、機械翻訳部 1 4 による機械翻訳後の文書が、第 N 言語の文書であるかどうかを判断する処理であってもよい。その場合には、機械翻訳後の文書が第 N 言語の文書である場合には、ステップ S 1 0 7 に進み、そうでなければ、ステップ S 1 0 3 に戻ることになる。また、図 2 のフローチャートにおいて、電源オフや処理終了の割り込みにより処理は終了する。

【 0 0 6 7 】

図 3 は、図 2 のフローチャートにおける選択された多言語対訳情報を用いた機械翻訳の

10

20

30

40

50

処理（ステップ S 1 0 4 の処理）の詳細を示すフローチャートである。


【 0 0 6 8 】

（ステップ S 2 0 1）機械翻訳手段 2 1 は、第 i 言語の翻訳対象文書を、第 $(i + 1)$ 言語の文書に翻訳する。その翻訳後の文書が翻訳結果文書である。

【 0 0 6 9 】


（ステップ S 2 0 2）翻訳結果文書蓄積手段 2 2 は、翻訳結果文書を記録媒体に蓄積する。

【 0 0 7 0 】

（ステップ S 2 0 3）翻訳ペア取得手段 2 4 は、第 i 言語の翻訳対象文書と、第 $(i + 1)$ 言語の翻訳結果文書とを用いて、翻訳ペアを取得する。この処理の詳細については、 図 4 のフローチャートを用いて後述する。


10

【 0 0 7 1 】



（ステップ S 2 0 4）置換ペア特定手段 2 5 は、翻訳ペア取得手段 2 4 によって取得された翻訳ペアから、その翻訳ペアの目的言語の単語が、選択後の多言語対訳情報に含まれる単語となる修正ペアを作成する。なお、翻訳ペアに含まれる両単語が、選択後の多言語対訳情報に含まれる場合には、置換ペア特定手段 2 5 は、その翻訳ペアをそのまま修正ペアとするものとする。なお、この処理の詳細については、 図 5 のフローチャートを用いて後述する。また、修正ペアは、第 i 言語の単語と、第 $(i + 1)$ 言語の単語とのペアであるが、翻訳ペアと同様に、そのうち、第 i 言語を原言語と呼び、第 $(i + 1)$ 言語を目的言語と呼ぶことにする。

20



【 0 0 7 2 】

（ステップ S 2 0 5）置換ペア特定手段 2 5 は、翻訳ペアと、修正ペアとを用いて、置換ペアを特定する。なお、この処理の詳細については、 図 6 のフローチャートを用いて後述する。

【 0 0 7 3 】

（ステップ S 2 0 6）翻訳結果文書修正手段 2 6 は、置換ペア特定手段 2 5 によって特定された置換ペアを用いて、翻訳結果文書を修正することによって、修正翻訳結果文書を作成する。そして、 図 2 のフローチャートに戻る。なお、この処理の詳細については、 図 7 のフローチャートを用いて後述する。

【 0 0 7 4 】


 図 4 は、 図 3 のフローチャートにおける翻訳ペアの取得の処理（ステップ S 2 0 3 の処理）の詳細を示すフローチャートである。

30

【 0 0 7 5 】

（ステップ S 3 0 1）翻訳ペア取得手段 2 4 は、カウンタ m を 1 に設定する。

【 0 0 7 6 】

（ステップ S 3 0 2）翻訳ペア取得手段 2 4 は、第 i 言語の翻訳対象文書の m 番目の単語を特定する。なお、この単語の特定の際に、翻訳ペア取得手段 2 4、あるいは、その他の構成要素によって、第 i 言語の翻訳対象文書の形態素解析が行われてもよい。日本語などの文書では、英語などの文書と異なり、単語の区切りが文書において明らかではないからである。他の文書において、単語の特定等がなされる場合にも同様であるとする。この特定された m 番目の単語は、 図示しない記録媒体等において一時的に保持されてもよく、あるいは、第 i 言語の翻訳対象文書において特定された単語にフラグ等が付加されてもよい。

40

【 0 0 7 7 】

（ステップ S 3 0 3）翻訳ペア取得手段 2 4 は、ステップ S 3 0 2 で特定した m 番目の単語が、選択後のいずれかの多言語対訳情報に含まれるかどうか判断する。この判断は、例えば、選択後の各多言語対訳情報に、特定した m 番目の単語が含まれるかどうか順番に判断することによってなされてもよい。そして、含まれる場合には、ステップ S 3 0 4 に進み、そうでない場合には、ステップ S 3 1 0 に進む。

【 0 0 7 8 】

50

(ステップ S 3 0 4) 翻訳ペア取得手段 2 4 は、対訳辞書記憶手段 2 3 で記憶されている、第 i 言語の単語と、第 $(i + 1)$ 言語の単語とを対応付ける対訳辞書を用いて、特定した m 番目の単語と対訳関係にある第 $(i + 1)$ 言語の 1 以上の単語を特定する。この特定された第 $(i + 1)$ 言語の 1 以上の単語は、図示しない記録媒体等において一時的に保持されてもよく、あるいは、対訳辞書において特定された単語にフラグ等が付加されてもよい。

【0079】

(ステップ S 3 0 5) 翻訳ペア取得手段 2 4 は、カウンタ n を 1 に設定する。

【0080】

(ステップ S 3 0 6) 翻訳ペア取得手段 2 4 は、ステップ S 3 0 4 で特定した第 $(i + 1)$ 言語の n 番目の単語が、機械翻訳手段 2 1 によって機械翻訳された第 $(i + 1)$ 言語の翻訳結果文書に含まれるかどうか判断する。そして、含まれる場合には、ステップ S 3 0 9 に進み、そうでない場合には、ステップ S 3 0 7 に進む。

10

【0081】

(ステップ S 3 0 7) 翻訳ペア取得手段 2 4 は、カウンタ n を 1 だけインクリメントする。

【0082】

(ステップ S 3 0 8) 翻訳ペア取得手段 2 4 は、ステップ S 3 0 4 で特定した第 $(i + 1)$ 言語の単語に、 n 番目の単語が存在するかどうか判断する。そして、存在する場合には、ステップ S 3 0 6 に戻り、そうでない場合には、ステップ S 3 1 0 に進む。

20

【0083】

(ステップ S 3 0 9) 翻訳ペア取得手段 2 4 は、第 i 言語の翻訳対象文書の m 番目の単語と、ステップ S 3 0 4 で特定した第 $(i + 1)$ 言語の n 番目の単語とをペアとして有する翻訳ペアを取得する。この翻訳ペアは、図示しない記録媒体等において一時的に保持されてもよく、あるいは、翻訳対象文書や翻訳結果文書において、その翻訳ペアに含まれる単語にフラグ等が付加されてもよい。

【0084】

(ステップ S 3 1 0) 翻訳ペア取得手段 2 4 は、カウンタ m を 1 だけインクリメントする。

【0085】

(ステップ S 3 1 1) 翻訳ペア取得手段 2 4 は、第 i 言語の翻訳対象文書に m 番目の単語が存在するかどうか判断する。そして、存在する場合には、ステップ S 3 0 2 に戻り、そうでない場合には、図 3 のフローチャートに戻る。

30

【0086】

図 5 は、図 3 のフローチャートにおける修正ペアの作成の処理 (ステップ S 2 0 4 の処理) の詳細を示すフローチャートである。

(ステップ S 4 0 1) 置換ペア特定手段 2 5 は、カウンタ m を 1 に設定する。

【0087】

(ステップ S 4 0 2) 置換ペア特定手段 2 5 は、翻訳ペア取得手段 2 4 によって取得された翻訳ペアのうち、 m 番目の翻訳ペアが、選択後のいずれかの多言語対訳情報に含まれるかどうか判断する。そして、含まれる場合には、ステップ S 4 0 5 に進み、含まれない場合には、ステップ S 4 0 3 に進む。

40

【0088】

(ステップ S 4 0 3) 置換ペア特定手段 2 5 は、 m 番目の翻訳ペアの原言語の単語を含む選択後の多言語対訳情報を特定し、その特定した選択後の多言語対訳情報に含まれる単語であって、 m 番目の翻訳ペアの目的言語と同じ言語である単語を特定する。なお、 m 番目の翻訳ペアの原言語の単語を含む選択後の多言語対訳情報が複数存在する場合には、置換ペア特定手段 2 5 は、前述のように、そのいずれかの多言語対訳情報を選択して、単語の特定を行うものとする。なお、 m 番目の翻訳ペアの原言語の単語を含む選択後の多言語対訳情報が存在しない場合には、ステップ S 4 0 6 に進んでもよい。

50

【0089】

(ステップS404)置換ペア特定手段25は、m番目の翻訳ペアの目的言語の単語を、ステップS403で特定した単語に置き換えたペアである修正ペアを作成する。この修正ペアは、図示しない記録媒体等において一時的に保持されてもよく、あるいは、翻訳対象文書や選択後の多言語対訳情報において、その修正ペアに含まれる単語にフラグ等が付加されてもよい。

【0090】

(ステップS405)置換ペア特定手段25は、m番目の翻訳ペアを修正ペアとする。この修正ペアは、図示しない記録媒体等において一時的に保持されてもよく、あるいは、翻訳対象文書や選択後の多言語対訳情報において、その修正ペアに含まれる単語にフラグ等が付加されてもよい。

10

【0091】

(ステップS406)置換ペア特定手段25は、カウンタmを1だけインクリメントする。

【0092】

(ステップS407)置換ペア特定手段25は、翻訳ペア取得手段24によって取得された翻訳ペアに、m番目の翻訳ペアが含まれるかどうか判断する。そして、含まれる場合には、ステップS402に戻り、そうでない場合には、図3のフローチャートに戻る。

【0093】

図6は、図3のフローチャートにおける置換ペアの作成の処理(ステップS205の処理)の詳細を示すフローチャートである。

20

【0094】

(ステップS501)置換ペア特定手段25は、カウンタmを1に設定する。

【0095】

(ステップS502)置換ペア特定手段25は、翻訳ペア取得手段24によって取得された翻訳ペアのうち、m番目の翻訳ペアが修正ペアの集合に含まれるかどうか判断する。置換ペア特定手段25は、例えば、m番目の翻訳ペアが、いずれかの修正ペアと一致する場合には、m番目の翻訳ペアが修正ペアの集合に含まれると判断してもよい。そして、含まれる場合には、ステップS508に進み、含まれない場合には、ステップS503に進む。

30

【0096】

(ステップS503)置換ペア特定手段25は、カウンタnを1に設定する。

【0097】

(ステップS504)置換ペア特定手段25は、m番目の翻訳ペアに含まれる原言語の単語と、n番目の修正ペアに含まれる原言語の単語とが一致するかどうか判断する。そして、一致する場合には、ステップS507に進み、そうでない場合には、ステップS505に進む。

【0098】

(ステップS505)置換ペア特定手段25は、カウンタnを1だけインクリメントする。

40

【0099】

(ステップS506)置換ペア特定手段25は、図5のフローチャートにおいて作成された修正ペアに、n番目の修正ペアが存在するかどうか判断する。そして、存在する場合には、ステップS504に戻り、そうでない場合には、ステップS508に進む。

【0100】

(ステップS507)置換ペア特定手段25は、m番目の翻訳ペアの目的言語である置換対象単語と、n番目の修正ペアの目的言語である置換結果単語とを含む置換ペアを特定する。この置換ペアは、図示しない記録媒体等において一時的に保持されてもよく、あるいは、翻訳結果文書や選択後の多言語対訳情報において、その置換ペアに含まれる単語にフラグ等が付加されてもよい。

50

【0101】

(ステップS508)置換ペア特定手段25は、カウンタmを1だけインクリメントする。

【0102】

(ステップS509)置換ペア特定手段25は、翻訳ペア取得手段24によって取得された翻訳ペアに、m番目の翻訳ペアが含まれるかどうか判断する。そして、含まれる場合には、ステップS502に戻り、そうでない場合には、図3のフローチャートに戻る。

【0103】

図7は、図3のフローチャートにおける翻訳結果文書の修正の処理(ステップS206の処理)の詳細を示すフローチャートである。

10

(ステップS601)翻訳結果文書修正手段26は、カウンタmを1に設定する。

【0104】

(ステップS602)翻訳結果文書修正手段26は、翻訳結果文書蓄積手段22が蓄積した第(i+1)言語の翻訳結果文書のm番目の単語を特定する。この特定されたm番目の単語は、図示しない記録媒体等において一時的に保持されてもよく、あるいは、第(i+1)言語の翻訳結果文書において特定された単語にフラグ等が付加されてもよい。

【0105】

(ステップS603)翻訳結果文書修正手段26は、カウンタnを1に設定する。

【0106】

(ステップS604)翻訳結果文書修正手段26は、ステップS602で特定した第(i+1)言語の翻訳結果文書のm番目の単語と、n番目の置換ペアに含まれる置換対象単語とが一致するかどうか判断する。そして、一致する場合には、ステップS607に進み、そうでない場合には、ステップS605に進む。

20

【0107】

(ステップS605)翻訳結果文書修正手段26は、カウンタnを1だけインクリメントする。

【0108】

(ステップS606)翻訳結果文書修正手段26は、n番目の置換ペアが存在するかどうか判断する。そして、存在する場合には、ステップS604に戻り、そうでない場合には、ステップS608に進む。

30

【0109】

(ステップS607)翻訳結果文書修正手段26は、第(i+1)言語の翻訳結果文書において、ステップS602で特定した第(i+1)言語の翻訳結果文書のm番目の単語を、n番目の置換ペアに含まれる置換対象単語に置換する。

【0110】

(ステップS608)翻訳結果文書修正手段26は、カウンタmを1だけインクリメントする。

【0111】

(ステップS609)翻訳結果文書修正手段26は、第(i+1)言語の翻訳結果文書にm番目の単語が存在するかどうか判断する。そして、存在する場合には、ステップS602に戻り、そうでない場合には、図3のフローチャートに戻る。なお、図7における一連の処理が終了した後の第(i+1)言語の翻訳結果文書、すなわち、適宜、単語の置換の行われた第(i+1)言語の翻訳結果文書が、修正翻訳結果文書となる。

40

【0112】

次に、本実施の形態による機械翻訳装置1の動作について、具体例を用いて説明する。この具体例において、機械翻訳装置1は、日英翻訳と、英独翻訳を行うことによって、日本語の翻訳対象文書をドイツ語に翻訳する場合について説明する。したがって、機械翻訳装置1は、第1言語から第3言語までの機械翻訳を行うことになり、N=3に設定されることになる。

【0113】

50

この具体例では、多言語対訳情報記憶部 12 において、互いに同義である、日本語の単語と、英語の単語と、ドイツ語の単語との組である多言語対訳情報が記憶されているものとする。図 8 は、この具体例で用いられる多言語対訳情報の一例を示す図である。図 8 において、各レコードが、日本語の単語、英語の単語、ドイツ語の単語を含む多言語対訳情報である。例えば、1 番目の多言語対訳情報は、日本語の単語「空(そら)」と、英語の単語「sky」と、ドイツ語の単語「Himmel」とを含んでいる。

【0114】

また、この具体例では、対訳辞書記憶手段 23 において、日英対訳辞書と、英独対訳辞書とが記憶されているものとする。図 9 は、この具体例で用いられる日英対訳辞書の一例を示す図である。図 9 で示されるように、日英対訳辞書は、原言語の単語と、目的言語の単語とを含む組を複数有する情報である。例えば、原言語(日本語)の単語「空(そら)」と、目的言語(英語)の単語「sky, air, heaven」とが対応付けられている。したがって、この日英対訳辞書を用いることによって、日本語の単語「空」から、「空」と対訳関係にある英語の単語「sky」「air」「heaven」を取得することができる。

10

【0115】

まず、機械翻訳装置 1 のユーザが、キーボードやマウス等の入力デバイスを用いることによって、翻訳対象文書「それが彼女の欠点だ」を機械翻訳装置 1 に入力したとする。すると、機械翻訳装置 1 の翻訳対象文書受付部 11 は、その翻訳対象文書を受け付け(ステップ S101)、機械翻訳部 14、及び多言語対訳情報選択部 13 に渡す。多言語対訳情報選択部 13 は、その日本語の翻訳対象文書を受け取ると、その翻訳対象文書に形態素解析を行って、単語ごとに分割する。そして、その分割後の単語を含む多言語対訳情報が選択される(ステップ S102, S103)。ここでは、翻訳対象文書に含まれる単語「それ」、「欠点」を含む多言語対訳情報が選択されたものとする。その選択された多言語対訳情報は、図示しない記録媒体で一時的に記憶される。

20

【0116】

機械翻訳部 14 は、多言語対訳情報選択部 13 によって選択された多言語対訳情報を用いて、日本語から英語への機械翻訳を行う(ステップ S104)。具体的には、機械翻訳手段 21 は、翻訳対象文書受付部 11 から受け取った翻訳対象文書「それが彼女の欠点だ」を、英語の翻訳結果文書「That is her fault.」に機械翻訳する(ステップ S201)。翻訳結果文書蓄積手段 22 は、その英語の翻訳結果文書を、図示しない記録媒体に蓄積する(ステップ S202)。

30

【0117】

また、翻訳ペア取得手段 24 は、選択後の多言語対訳情報と、翻訳対象文書受付部 11 から受け取った日本語の翻訳対象文書「それが彼女の欠点だ」と、翻訳結果文書蓄積手段 22 が蓄積した英語の翻訳結果文書「That is her fault.」とを用いて、翻訳ペアを取得する(ステップ S203)。

【0118】

具体的には、翻訳ペア取得手段 24 は、翻訳対象文書の 1 番目の単語「それ」を特定し(ステップ S301, S302)、その単語「それ」が選択後の多言語対訳情報に含まれるかどうか判断する(ステップ S303)。翻訳ペア取得手段 24 は、例えば、その単語「それ」を検索キーとして選択後の多言語対訳情報を検索し、その検索がヒットした場合には、その単語が選択後の多言語対訳情報に含まれると判断する。この場合には、前述のように、単語「それ」を含む多言語対訳情報が選択されているため、翻訳ペア取得手段 24 は、単語「それ」が選択後の多言語対訳情報に含まれると判断する。そして、翻訳ペア取得手段 24 は、対訳辞書記憶手段 23 で記憶されている日英対訳辞書を用いて、その単語「それ」に対応する英語の 1 以上の単語「it」「that」を特定する(ステップ S304)。翻訳ペア取得手段 24 は、特定した英語の単語のうちの 1 番目の単語「it」が英語の翻訳結果文書に含まれるかどうか判断する(ステップ S305, S306)。例えば、翻訳ペア取得手段 24 は、単語「it」を検索キーとして翻訳結果文書を検索し、

40

50

その検索がヒットした場合には、その単語「i t」が翻訳結果文書に含まれると判断する。この場合には含まれないため、翻訳ペア取得手段24は、次の単語「t h a t」について、同様の処理を行う(ステップS307, S308, S306)。この場合には、単語「t h a t」が翻訳結果文書に含まれるため、翻訳ペア取得手段24は、原言語の単語「それ」と、目的言語の単語「t h a t」とを含む翻訳ペアを生成し、図示しない記録媒体に蓄積する(ステップS309)。図10の1番目のレコードは、そのようにして蓄積された翻訳ペアである。

【0119】

次に、翻訳ペア取得手段24は、日本語の翻訳対象文書の2番目の単語「が」を特定し(ステップS310, S311, S302)、その単語「が」が選択後の多言語対訳情報に含まれるかどうか判断する(ステップS303)。この場合には、その単語「が」が選択後の多言語対訳情報に含まれなかったとする。すると、翻訳ペア取得手段24は、次の単語について、同様の処理を繰り返す(ステップS310, S311, S302)。このような処理が繰り返して実行されることにより、日本語の翻訳対象文書と、英語の翻訳結果文書とを用いた翻訳ペアの取得の処理が終了したとする。そして、図示しない記録媒体では、図10で示されるように、2個の翻訳ペアが一時的に記憶されているものとする。

10

【0120】

次に、置換ペア特定手段25は、選択後の多言語対訳情報と、翻訳ペア取得手段24が取得した図10で示される翻訳ペアとを用いて、修正ペアを作成する(ステップS204)。具体的には、置換ペア特定手段25は、図10で示される1番目の翻訳ペアが、選択後の多言語対訳情報に含まれるかどうか判断する(ステップS401, S402)。置換ペア特定手段25は、例えば、その1番目の翻訳ペアの原言語の単語「それ」と、目的言語の単語「t h a t」とを検索キーとして、選択後の多言語対訳情報を検索し、両方の単語を含む一の多言語対訳情報が存在するかどうか判断する。この場合には、存在したとする。すると、置換ペア特定手段25は、その翻訳ペアを修正ペアとして図示しない記録媒体に蓄積する(ステップS405)。図11の1番目のレコードは、そのようにして蓄積された修正ペアである。その後、置換ペア特定手段25は、2番目の翻訳ペアについても、同様にして、選択後の多言語対訳情報に含まれるかどうか判断する(ステップS406, S407, S402)。この場合にも、2番目の翻訳ペアが選択後の多言語対訳情報に含まれるとすると、置換ペア特定手段25は、その翻訳ペアを修正ペアとして図示しない記録媒体に蓄積する(ステップS405)。そして、修正ペアを作成する処理は終了となる(ステップS406, S407)。図11は、そのようにして作成された修正ペアを示す図である。図10, 図11で示されるように、日英翻訳の場合には、翻訳ペアと修正ペアとがまったく同じになる。

20

30

【0121】

次に、置換ペア特定手段25は、図10で示される翻訳ペアと、図11で示される修正ペアとを用いて、置換ペアを特定する処理を行う(ステップS205)。この場合には、すべての翻訳ペアは、修正ペアの集合に含まれると判断され(ステップS501, S502, S508, S509)、置換ペア特定手段25による置換ペアの特定は行われない。したがって、翻訳結果文書の修正の処理(ステップS206)においても、置換ペアが存在しないため、翻訳結果文書の単語が置換ペアの置換対象単語と等しいと判断されることはなく、翻訳結果文書修正手段26による翻訳結果文書の修正は行われない(ステップS601~S606, S608, S609)。その結果、翻訳結果文書蓄積手段22が蓄積した英語の翻訳結果文書は修正されず、その翻訳結果文書そのものが、英独機械翻訳の翻訳対象文書となる。

40

【0122】

その後、再度、多言語対訳情報選択部13による多言語対訳情報の選択が行われる(ステップS105, S106, S103)。この場合には、多言語対訳情報選択部13は、英語の翻訳対象文書「T h a t i s h e r f a u l t .」に含まれる単語を含む多言語対訳情報を選択することになる。その選択後の多言語対訳情報は、図12で示される

50

ようになったとする。その選択された多言語対訳情報は、図示しない記録媒体で記憶される。

【0123】

次に、機械翻訳部14は、多言語対訳情報選択部13によって選択された多言語対訳情報を用いて、英語からドイツ語への機械翻訳を行う(ステップS104)。具体的には、機械翻訳手段21は、翻訳結果文書蓄積手段22が翻訳結果文書を蓄積した記録媒体から、英語の翻訳対象文書「That is her fault.」を読み出し、ドイツ語の翻訳結果文書「Das ist ihre Schuld.」に機械翻訳する(ステップS201)。翻訳結果文書蓄積手段22は、その英語の翻訳結果文書を、図示しない記録媒体に蓄積する(ステップS202)。

10

【0124】

翻訳ペア取得手段24は、選択後の多言語対訳情報と、英語の翻訳対象文書「That is her fault.」と、ドイツ語の翻訳結果文書「Das ist ihre Schuld.」とを用いて、翻訳ペアを取得する(ステップS203)。この処理は、前述の説明と同様であり、詳細な説明を省略する。この翻訳ペアを取得する処理の結果、図13で示される2個の翻訳ペアが取得されたとする。

【0125】

次に、置換ペア特定手段25は、選択後の多言語対訳情報と、翻訳ペア取得手段24が取得した図13で示される翻訳ペアとを用いて、修正ペアを作成する(ステップS204)。具体的には、置換ペア特定手段25は、図13で示される1番目の翻訳ペアが、図12で示される選択後の多言語対訳情報に含まれるかどうか判断する(ステップS401, S402)。この場合には、その翻訳ペアが選択後の多言語対訳情報に含まれるため、置換ペア特定手段25は、その翻訳ペアを修正ペアとして図示しない記録媒体に蓄積する(ステップS405)。図14の1番目のレコードは、そのようにして蓄積された修正ペアである。その後、置換ペア特定手段25は、図13で示される2番目の翻訳ペアについても、同様にして、選択後の多言語対訳情報に含まれるかどうか判断する(ステップS406, S407, S402)。この場合には、2番目の翻訳ペアが、図12で示される選択後の多言語対訳情報に含まれないため、置換ペア特定手段25は、その翻訳ペアに含まれる原言語の単語「fault」を含む選択後の多言語対訳情報に含まれる目的言語(ここではドイツ語)の単語を特定する(ステップS403)。例えば、置換ペア特定手段25は、図12で示される選択後の多言語対訳情報において、翻訳ペアに含まれる原言語の単語「fault」を検索キーとして検索を行い、ヒットした多言語対訳情報を特定する。そして、その特定した多言語対訳情報に含まれる、翻訳ペアの目的言語であるドイツ語の単語を特定する。この場合には、ドイツ語の2個の単語「Fehler」「Mangle」が特定される。2個の単語が特定されたため、置換ペア特定手段25は、前述のように、何らかの方法によって、一方の単語を選択する。ここでは、「Fehler」が選択されたとする。すると、置換ペア特定手段25は、翻訳ペアに含まれる原言語の単語「fault」と、選択されたドイツ語の単語「Fehler」とを含む修正ペアを作成し、図示しない記録媒体に蓄積する(ステップS404)。そして、修正ペアを作成する処理は終了となる(ステップS406, S407)。図14は、そのようにして作成された修正ペアを示す図である。

20

30

40

【0126】

次に、置換ペア特定手段25は、図13で示される翻訳ペアと、図14で示される修正ペアとを用いて、置換ペアを特定する処理を行う(ステップS205)。具体的には、置換ペア特定手段25は、図13で示される1番目の翻訳ペアが、図13で示される修正ペアの集合に含まれるかどうか判断する(ステップS501, S502)。この場合には、翻訳ペア(that, das)が修正ペアの集合に含まれるため、置換ペア特定手段25は、次の翻訳ペアについて、同様の判断を行う(ステップS508, S509, S502)。この場合には、翻訳ペア(fault, Schuld)は、図14で示される修正ペアの集合に含まれないため、置換ペア特定手段25は、翻訳ペアの原言語の単語「fau

50

l t」が、1番目の修正ペアの原言語の単語「t h a t」と等しいかどうか判断する（ステップS 5 0 3，S 5 0 4）。この場合には、両者は等しくないため、置換ペア特定手段25は、次の修正ペアについて、同様の判断を行う（ステップS 5 0 5，S 5 0 6，S 5 0 4）。この場合には、翻訳ペアの原言語の単語「f a u l t」と、2番目の修正ペアの原言語の単語「f a u l t」とが等しいため、置換ペア特定手段25は、翻訳ペアの原言語の単語「f a u l t」である置換対象単語と、2番目の修正ペアの目的言語の単語「F e h l e r」である置換結果単語とを含む置換ペアを特定し、図示しない記録媒体に蓄積する（ステップS 5 0 7）。そして、置換ペアを作成する処理は終了となる（ステップS 5 0 8，S 5 0 9）。図15は、そのようにして作成された修正ペアを示す図である。

【0127】

次に、翻訳結果文書修正手段26は、置換ペア特定手段25が特定した図15で示される置換ペアと、翻訳結果文書蓄積手段22が蓄積したドイツ語の翻訳結果文書「D a s i s t i h r e S c h u l d .」とを用いて、翻訳結果文書の修正を行う（ステップS 2 0 6）。具体的には、翻訳結果文書修正手段26は、ドイツ語の翻訳結果文書「D a s i s t i h r e S c h u l d .」の1番目の単語「D a s」を特定し（ステップS 6 0 1，S 6 0 2）、その単語が置換ペアの置換対象単語と一致するかどうか判断する（ステップS 6 0 3，S 6 0 4）。この場合には、両単語は一致せず、また、2番目以降の置換ペアも存在しないため（ステップS 6 0 5，S 6 0 6）、翻訳結果文書修正手段26は、次の単語「i s t」についても同様の処理を行う（ステップS 6 0 8，S 6 0 9，S 6 0 2～S 6 0 4）。この場合にも、その単語と置換ペアの置換対象単語とは一致せず、他の置換ペアも存在しないため（ステップS 6 0 5，S 6 0 6）、翻訳結果文書修正手段26は、次の単語「i h r e」についても同様の処理を行う（ステップS 6 0 8，S 6 0 9，S 6 0 2～S 6 0 4）。この場合にも、その単語と置換ペアの置換対象単語とは一致せず、他の置換ペアも存在しないため（ステップS 6 0 5，S 6 0 6）、翻訳結果文書修正手段26は、次の単語「S c h u l d」についても同様の処理を行う（ステップS 6 0 8，S 6 0 9，S 6 0 2～S 6 0 4）。この場合には、その単語と置換ペアの置換対象単語とが一致するため、翻訳結果文書修正手段26は、その単語「S c h u l d」を、置換ペアの置換結果単語「F e h l e r」に置換する（ステップS 6 0 7）。そして、それ以上の単語が翻訳結果文書に含まれないため、翻訳結果文書を修正する処理は終了となる（ステップS 6 0 8，S 6 0 9）。その結果、翻訳結果文書修正手段26によって修正された後の修正翻訳結果文書は、「D a s i s t i h r e F e h l e r .」となる。

【0128】

その後、出力部15は、そのドイツ語の修正翻訳結果文書「D a s i s t i h r e F e h l e r .」を出力する（ステップS 1 0 5～S 1 0 7）。出力部15が、例えば、図示しないディスプレイに修正翻訳結果文書「D a s i s t i h r e F e h l e r .」を表示した場合には、ユーザは、その表示を見ることによって、入力した日本語の文書「それが彼女の欠点だ」に対応するドイツ語の翻訳結果を知ることができる。

【0129】

ここで、本実施の形態による機械翻訳装置1の評価について説明する。本実施の形態による機械翻訳装置1と、従来の機械翻訳装置、すなわち、二言語間での機械翻訳を単に繰り返すだけの装置とを用いて機械翻訳を行った。評価用の例文としては、N T Tが提供している機械翻訳性能評価用例文を元にした100文を用いた。機械翻訳は、日英翻訳、英独翻訳、独英翻訳、英日翻訳の4個の二言語間の翻訳を連携した日独折り返し翻訳を行った。評価者は3人であり、評価値は5段階である。

【0130】

図16は、従来の機械翻訳装置を用いた場合の評価値の平均と、本実施の形態による機械翻訳装置1を用いた場合の評価値の平均とを比較する表である。なお、図16において、本実施の形態による機械翻訳装置1を用いた場合を「適用後」と示している。図16から分かるように、3人の評価者すべてにおいて、本実施の形態による機械翻訳装置1を用いることによって、従来例の場合よりも評価値の平均が上がっている。このことは、本実

10

20

30

40

50

施の形態による機械翻訳装置 1 を用いることによって、訳語のドリフトの発生を抑制することができ、機械翻訳の精度がより高くなっているからであると考えられる。

【0131】

図 17 は、従来例の評価値別に、評価値が上がった文の割合を示す表である。例えば、従来の機械翻訳装置を用いた場合の評価値が「3」であった文のうち、平均して 32% の文の評価値が、本実施の形態による機械翻訳装置 1 を用いることによって改善している、すなわち、評価値が 4 または 5 になっていることが分かる。したがって、本実施の形態による機械翻訳装置 1 を用いることによって、平均して 3 割から 6 割程度の文の評価値の向上に寄与できることが分かる。

【0132】

以上のように、本実施の形態による機械翻訳装置 1 によれば、二言語間の機械翻訳を繰り返すことによって第 1 言語から第 N 言語までの翻訳を行う際に、多言語対訳情報を用いることによって、訳語のドリフトの発生を抑制することができる。したがって、第 N 言語の翻訳結果が、第 1 言語の翻訳対象と同じ意味の文書になるようにすることができる。

【0133】

なお、本実施の形態では、翻訳ペアを取得し、その翻訳ペアを用いて修正ペアを作成し、その修正ペアを用いて置換ペアを特定する場合について説明したが、修正ペアを作成することなく、置換ペアを特定してもよい。例えば、置換ペア特定手段 25 は、ある翻訳ペアが選択後の多言語対訳情報に含まれない場合に、その翻訳ペアの原言語の単語を含む選択後の多言語対訳情報に含まれる目的言語（翻訳ペアの目的言語である）の単語を置換結果単語とし、その翻訳ペアの原言語の単語を置換対象単語とする置換ペアを特定してもよい。このように、翻訳ペアの取得から置換ペアの特定に至るまでの処理は、各種の処理が存在しうることになり、それらは限定されるものではない。

【0134】

また、本実施の形態では、置換ペア特定手段 25 が、2 以上の置換ペアを特定することができたとしても、一の置換ペアのみを特定する場合について説明したが、置換ペア特定手段 25 は、複数の置換ペアを特定してもよい。その場合には、翻訳結果文書修正手段 26 は、いずれかの置換ペアを用いて一の修正翻訳結果文書を生成してもよく、あるいは、複数の修正翻訳結果文書を生成してもよい。後者の場合には、その各修正翻訳結果文書を翻訳対象文書として、その後の機械翻訳等が行われていくことになる。その結果として、最終的に第 N 言語の複数の修正翻訳結果文書が生成されることになる。その後、出力部 15 は、それらすべてを出力してもよく、あるいは、第 N 言語の複数の修正翻訳結果文書から選択された一の文書を出力してもよい。複数の修正翻訳結果文書から、出力する一の文書を選択する方法としては、例えば、その文書を生成する際に実行された単語の置換回数が最も少ないものを選択する方法等がある。なお、このように複数の修正翻訳結果情報を扱う場合には、多言語対訳情報選択部 13 による選択は、翻訳対象文書ごと、すなわち、修正翻訳結果文書ごとになされることになる。したがって、翻訳対象文書ごとに、選択後の多言語対訳情報が管理されることになる。

【0135】

また、本実施の形態では、機械翻訳部 14 が機械翻訳手段 21 や翻訳結果文書蓄積手段 22 等を備える場合について説明したが、前述のように、機械翻訳部 14 は、選択後の多言語対訳情報を取り込み、機械翻訳の仕組み自体を変更することによって、その選択後の多言語対訳情報を用いた機械翻訳がなされるようにしてもよい。その場合には、機械翻訳部 14 では、置換ペアの特定や、置換ペアを用いた翻訳結果文書の修正等の処理が行われないことになる。なお、この場合に、多言語対訳情報の選択の処理（ステップ S103 の処理）を行うために、第 i 言語の翻訳対象文書に含まれる原言語単語と、第 (i + 1) 言語の翻訳結果文書に含まれる目的言語の単語とのペアであって、互いに対訳関係にある単語のペアを特定し、その特定したペアを用いて多言語対訳情報の選択を行ってもよく、そうでなくてもよい。前者の場合に、そのペアを特定する方法は、例えば、機械翻訳部 14 が機械翻訳で用いた原言語の単語と、目的言語の単語とのペアを機械翻訳部 14 から受け

10

20

30

40

50

取る方法であってもよく、あるいは、翻訳ペアを取得する方法と同様の方法であってもよい。そのペアを用いて多言語対訳情報の選択を行う場合には、多言語対訳情報選択部 13 は、その単語のペアの有する両単語を含む多言語対訳情報を選択することになる。

【0136】

また、上記実施の形態では、機械翻訳装置がスタンドアロンである場合について説明したが、機械翻訳装置は、スタンドアロンの装置であってもよく、サーバ・クライアントシステムにおけるサーバ装置であってもよい。後者の場合には、出力部や受付部は、通信回線を介して入力を受け付けたり、画面を出力したりすることになる。

【0137】

また、上記実施の形態において、各処理または各機能は、単一の装置または単一のシステムによって集中処理されることによって実現されてもよく、あるいは、複数の装置または複数のシステムによって分散処理されることによって実現されてもよい。

10

【0138】

また、上記実施の形態において、各構成要素が実行する処理に関する情報、例えば、各構成要素が受け付けたり、取得したり、選択したり、生成したり、送信したり、受信したりした情報や、各構成要素が処理で用いるしきい値や数式、アドレス等の情報等は、上記説明で明記していない場合であっても、図示しない記録媒体において、一時的に、あるいは長期にわたって保持されていてもよい。また、その図示しない記録媒体への情報の蓄積を、各構成要素、あるいは、図示しない蓄積部が行ってもよい。また、その図示しない記録媒体からの情報の読み出しを、各構成要素、あるいは、図示しない読み出し部が行ってもよい。

20

【0139】

また、上記実施の形態において、各構成要素等で用いられる情報、例えば、各構成要素が処理で用いるしきい値やアドレス、各種の設定値等の情報がユーザによって変更されてもよい場合には、上記説明で明記していない場合であっても、ユーザが適宜、それらの情報を変更できるようにしてもよく、あるいは、そうでなくてもよい。それらの情報をユーザが変更可能な場合には、その変更は、例えば、ユーザからの変更指示を受け付ける図示しない受付部と、その変更指示に応じて情報を変更する図示しない変更部とによって実現されてもよい。その図示しない受付部による変更指示の受け付けは、例えば、入力デバイスからの受け付けでもよく、通信回線を介して送信された情報の受信でもよく、所定の記録媒体から読み出された情報の受け付けでもよい。

30

【0140】

また、上記実施の形態において、機械翻訳装置に含まれる 2 以上の構成要素が通信デバイスや入力デバイス等を有する場合に、2 以上の構成要素が物理的に単一のデバイスを有してもよく、あるいは、別々のデバイスを有してもよい。

【0141】

また、上記実施の形態において、各構成要素は専用のハードウェアにより構成されてもよく、あるいは、ソフトウェアにより実現可能な構成要素については、プログラムを実行することによって実現されてもよい。例えば、ハードディスクや半導体メモリ等の記録媒体に記録されたソフトウェア・プログラムを CPU 等のプログラム実行部が読み出して実行することによって、各構成要素が実現され得る。なお、上記実施の形態における機械翻訳装置を実現するソフトウェアは、以下のようなプログラムである。つまり、このプログラムは、コンピュータを、二言語間の機械翻訳を繰り返すことによって、第 1 言語から第 N 言語 (N は 3 以上の整数である) までの翻訳を行う機械翻訳装置として機能させるためのプログラムであって、第 1 言語の翻訳対象となる文書である翻訳対象文書を受け付ける翻訳対象文書受付部と、第 1 言語から第 N 言語までの同義の単語の組である多言語対訳情報が 1 以上記憶される多言語対訳情報記憶部で記憶されている 1 以上の多言語対訳情報から、第 i 言語 (i は 1 から N - 1 の整数である) の翻訳対象文書に含まれる単語を含む多言語対訳情報を選択する多言語対訳情報選択部と、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる二言語間の対訳関係が用いられるように、前記翻訳対象文書受

40

50

付部が受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 ($i + 1$) 言語に機械翻訳する処理を第 N 言語への機械翻訳を行うまで繰り返す機械翻訳部と、前記機械翻訳部が機械翻訳した第 N 言語の文書を出力する出力部として機能させるためのものである。

【 0 1 4 2 】

また、このプログラムにおいて、前記機械翻訳部は、前記翻訳対象文書受付部が受け付けた第 1 言語の翻訳対象文書からはじめて、第 i 言語の翻訳対象文書を第 ($i + 1$) 言語に機械翻訳する処理を繰り返す機械翻訳手段と、前記機械翻訳手段が機械翻訳した結果の文書である翻訳結果文書を蓄積する翻訳結果文書蓄積手段と、翻訳対象文書に含まれる単語と、当該翻訳対象文書を前記機械翻訳手段が機械翻訳した翻訳結果文書に含まれる単語とのペアであって、対訳関係にある単語のペアである翻訳ペアを取得する翻訳ペア取得手段と、前記翻訳ペア取得手段が取得した翻訳ペアのうち、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれない翻訳ペアに含まれる目的言語の単語である置換対象単語と、当該翻訳ペアに含まれる原言語の単語を含む、前記多言語対訳情報選択部が選択した多言語対訳情報に含まれる前記目的言語の単語である置換結果単語とのペアである置換ペアを特定する置換ペア特定手段と、前記翻訳結果文書蓄積手段が蓄積した翻訳結果文書に含まれる単語のうち、前記置換ペア特定手段が特定した置換ペアに含まれる置換対象単語を、当該置換ペアに含まれる置換結果単語に置換した文書である修正翻訳結果文書を作成する翻訳結果文書修正手段と、を備え、前記機械翻訳手段は、前記翻訳結果文書修正手段が生成した修正翻訳結果文書を翻訳対象文書として機械翻訳し、前記出力部は、前記翻訳結果文書修正手段が生成した第 N 言語の修正翻訳結果文書を出力してもよい。

10

20

【 0 1 4 3 】

なお、上記プログラムにおいて、上記プログラムが実現する機能には、ハードウェアでしか実現できない機能は含まれない。例えば、情報を受け付ける受付部や、情報を出力する出力部などにおけるモデムやインターフェースカードなどのハードウェアでしか実現できない機能は、上記プログラムが実現する機能には少なくとも含まれない。

【 0 1 4 4 】

また、このプログラムは、サーバなどからダウンロードされることによって実行されてもよく、所定の記録媒体（例えば、CD-ROMなどの光ディスクや磁気ディスク、半導体メモリなど）に記録されたプログラムが読み出されることによって実行されてもよい。また、そのプログラムが製品に取り込まれ、その製品で用いられることによって、プログラムプロダクトが構成されてもよい。

30

【 0 1 4 5 】

また、このプログラムを実行するコンピュータは、単数であってもよく、複数であってもよい。すなわち、集中処理を行ってもよく、あるいは分散処理を行ってもよい。

【 0 1 4 6 】

図 18 は、上記プログラムを実行して、上記実施の形態による機械翻訳装置 1 を実現するコンピュータの外観の一例を示す模式図である。上記実施の形態は、コンピュータハードウェア及びその上で実行されるコンピュータプログラムによって実現される。

40

【 0 1 4 7 】

図 18 において、コンピュータシステム 100 は、CD-ROM (Compact Disk Read Only Memory) ドライブ 105、FD (Flexible Disk) ドライブ 106 を含むコンピュータ 101 と、キーボード 102 と、マウス 103 と、モニタ 104 とを備える。

【 0 1 4 8 】

図 19 は、コンピュータシステムを示す図である。図 19 において、コンピュータ 101 は、CD-ROM ドライブ 105、FD ドライブ 106 に加えて、CPU (Central Processing Unit) 111 と、ブートアッププログラム等のプログラムを記憶するための ROM (Read Only Memory) 112 と、CPU 111 に接続され、アプリケーションプログラムの命令を一時的に記憶すると共に、一時記

50

憶空間を提供するRAM (Random Access Memory) 113と、アプリケーションプログラム、システムプログラム、及びデータを記憶するハードディスク114と、CPU111、ROM112等を相互に接続するバス115とを備える。なお、コンピュータ101は、LANへの接続を提供する図示しないネットワークカードを含んでいてもよい。

【0149】

コンピュータシステム100に、上記実施の形態による機械翻訳装置1の機能を実行させるプログラムは、CD-ROM121、またはFD122に記憶されて、CD-ROMドライブ105、またはFDドライブ106に挿入され、ハードディスク114に転送されてもよい。これに代えて、そのプログラムは、図示しないネットワークを介してコンピュータ101に送信され、ハードディスク114に記憶されてもよい。プログラムは実行の際にRAM113にロードされる。なお、プログラムは、CD-ROM121やFD122、またはネットワークから直接、ロードされてもよい。

10

【0150】

プログラムは、コンピュータ101に、上記実施の形態による機械翻訳装置1の機能を実行させるオペレーティングシステム(OS)、またはサードパーティプログラム等を必ずしも含んでいなくてもよい。プログラムは、制御された態様で適切な機能(モジュール)を呼び出し、所望の結果が得られるようにする命令の部分のみを含んでいてもよい。コンピュータシステム100がどのように動作するのかについては周知であり、詳細な説明は省略する。

20

【0151】

また、本発明は、以上の実施の形態に限定されることなく、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

【産業上の利用可能性】

【0152】

以上より、本発明による機械翻訳装置等によれば、二言語間の機械翻訳を繰り返すことによって第1言語から第N言語(Nは3以上の整数)までの翻訳を行う際に、多言語対訳情報を用いることによって、訳語のドリフトの発生を抑制することができるという効果が得られ、機械翻訳を行う装置等として有用である。

30

【図面の簡単な説明】

【0153】

【図1】本発明の実施の形態1による機械翻訳装置の構成を示すブロック図

【図2】同実施の形態による機械翻訳装置の動作を示すフローチャート

【図3】同実施の形態による機械翻訳装置の動作を示すフローチャート

【図4】同実施の形態による機械翻訳装置の動作を示すフローチャート

【図5】同実施の形態による機械翻訳装置の動作を示すフローチャート

【図6】同実施の形態による機械翻訳装置の動作を示すフローチャート

【図7】同実施の形態による機械翻訳装置の動作を示すフローチャート

【図8】同実施の形態における多言語対訳情報の一例を示す図

【図9】同実施の形態における日英対訳辞書の一例を示す図

40

【図10】同実施の形態における翻訳ペアの一例を示す図

【図11】同実施の形態における修正ペアの一例を示す図

【図12】同実施の形態における選択後の多言語対訳情報の一例を示す図

【図13】同実施の形態における翻訳ペアの一例を示す図

【図14】同実施の形態における修正ペアの一例を示す図

【図15】同実施の形態における置換ペアの一例を示す図

【図16】同実施の形態における評価結果の従来例との比較を示す図

【図17】同実施の形態における評価値の改善割合を示す図

【図18】同実施の形態におけるコンピュータシステムの外観一例を示す模式図

【図19】同実施の形態におけるコンピュータシステムの構成の一例を示す図

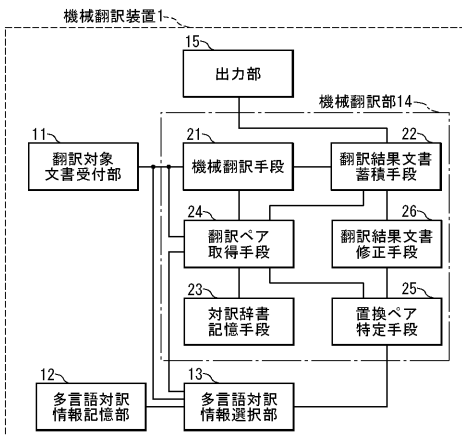
50

【符号の説明】

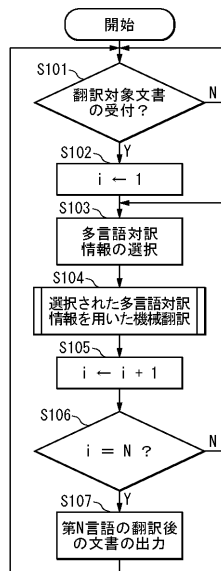
【0154】

- 1 機械翻訳装置
- 11 翻訳対象文書受付部
- 12 多言語対訳情報記憶部
- 13 多言語対訳情報選択部
- 14 機械翻訳部
- 15 出力部
- 21 機械翻訳手段
- 22 翻訳結果文書蓄積手段
- 23 対訳辞書記憶手段
- 24 翻訳ペア取得手段
- 25 置換ペア特定手段
- 26 翻訳結果文書修正手段

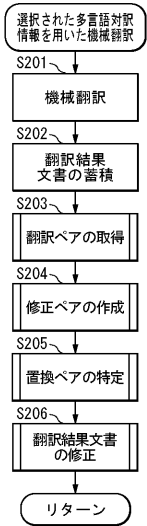
【図1】



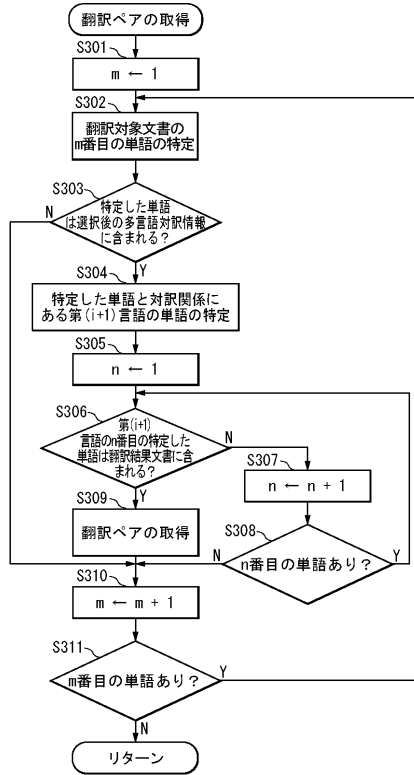
【図2】



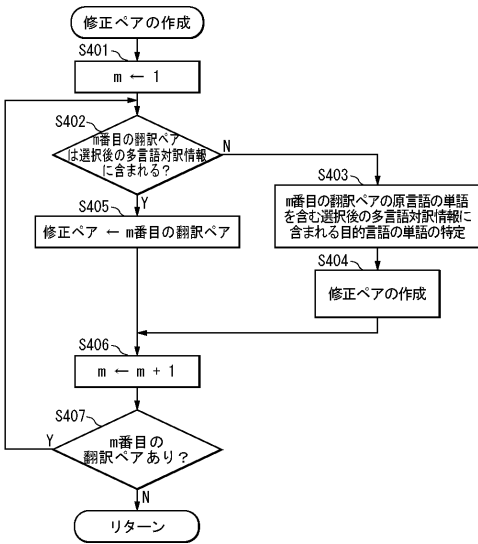
【 図 3 】



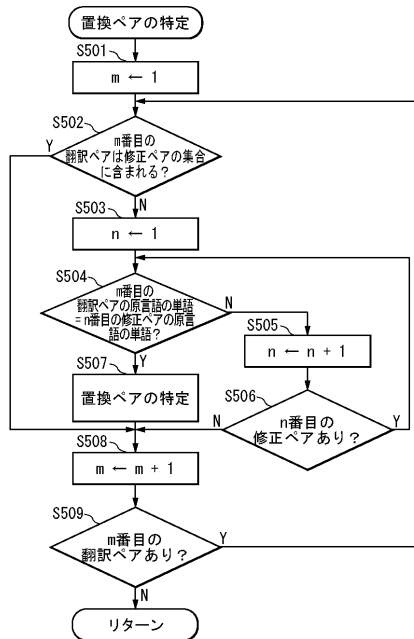
【 図 4 】



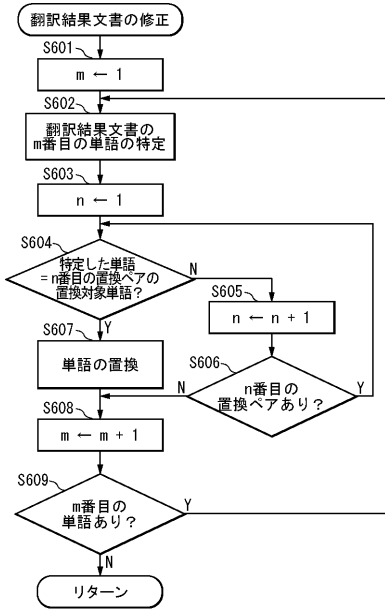
【 図 5 】



【 図 6 】



【 図 7 】



【 図 8 】

多言語対訳情報	
空	sky, Himmel
海	sea, Meer
⋮	⋮
欠点	fault, Fehler
欠点	fault, Mangel
⋮	⋮

【 図 9 】

日英対訳辞書

原言語の単語	目的言語の単語
空	sky, air, heaven
欠点	fault, defect, weakness
それ	it, that
⋮	⋮

【 図 10 】

翻訳ペア
(それ, that)
(欠点, fault)

【 図 11 】

修正ペア
(それ, that)
(欠点, fault)

【 図 15 】

置換ペア
(Schuld, Fehler)

【 図 12 】

選択後の多言語対訳情報
それ, that, das
欠点, fault, Fehler
欠点, fault, Mangel

【 図 16 】

	評価者A	評価者B	評価者C	平均
従来例の評価値平均	2.34	2.58	2.33	2.41667
適用後の評価値平均	2.8	3.15	2.72	2.89
変化	+0.46	+0.57	+0.39	+0.473333

【 図 13 】

翻訳ペア
(that, das)
(fault, Schuld)

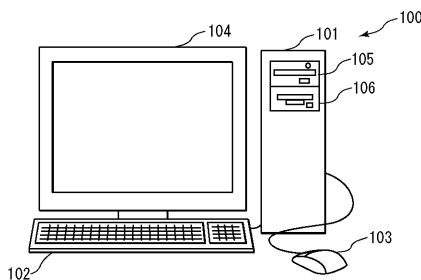
【 図 17 】

従来例の評価値	評価者A	評価者B	評価者C	平均
4	37.5	38.46154	25	33.65385
3	21.42857	37.14286	38.46154	32.34432
2	61.11111	58.62069	28	49.24393
1	47.72727	88.88889	42.42424	59.68013

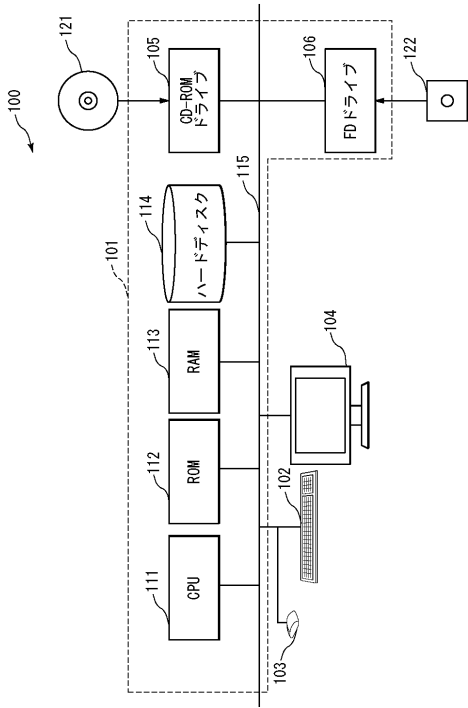
【 図 14 】

修正ペア
(that, das)
(fault, Fehler)

【 図 18 】



【図 19】



フロントページの続き

(72)発明者 村上 陽平

東京都小金井市貫井北町4 - 2 - 1 独立行政法人情報通信研究機構内

Fターム(参考) 5B091 AA01 BA11 CA21 CC01