

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第4869281号
(P4869281)

(45) 発行日 平成24年2月8日 (2012.2.8)

(24) 登録日 平成23年11月25日 (2011.11.25)

(51) Int.Cl.
G 0 6 F 17/28 (2006.01)

F I
G O 6 F 17/28 T

請求項の数 7 (全 15 頁)

(21) 出願番号	特願2008-102395 (P2008-102395)	(73) 特許権者	000003078
(22) 出願日	平成20年4月10日 (2008.4.10)		株式会社東芝
(65) 公開番号	特開2009-252143 (P2009-252143A)		東京都港区芝浦一丁目1番1号
(43) 公開日	平成21年10月29日 (2009.10.29)	(73) 特許権者	301063496
審査請求日	平成20年4月10日 (2008.4.10)		東芝ソリューション株式会社
			東京都港区芝浦一丁目1番1号
		(74) 代理人	100100516
			弁理士 三谷 恵
		(72) 発明者	佐々木 晶
			東京都港区芝浦一丁目1番1号 東芝ソリ ューション株式会社内
		審査官	成瀬 博之
		最終頁に続く	

(54) 【発明の名称】 機械翻訳装置、プログラム及び方法

(57) 【特許請求の範囲】

【請求項1】

翻訳に必要な知識情報・規則を蓄積した翻訳辞書部を記憶した記憶装置を備えた機械翻訳装置において、前記機械翻訳装置は、入力装置から入力された第1言語の原文に対してユーザから指定された第1言語の語句を受け付ける主要キーワード指定部と、入力された原文全体に対して、前記主要キーワード指定部で指定された第1言語の語句またはその語句を含むフレーズの訳語が、翻訳後の第2言語の訳文の文頭に最も近い位置に存在する訳文を生成する規則を選択して翻訳を行う翻訳部と、翻訳対象の原文や前記翻訳部による翻訳後の第2言語の訳文を出力装置に出力処理する出力処理部とを備える機械翻訳装置。

【請求項2】

前記入力装置からの指令に基づき入力された前記原文からキーワードを推定するキーワード推定部を設け、前記出力処理部は前記原文とともに前記キーワード推定部で推定されたキーワードを前記出力装置に出力する請求項1記載の機械翻訳装置。

【請求項3】

翻訳に必要な知識情報・規則を蓄積した翻訳辞書部を記憶した記憶装置を備えた機械翻訳装置において、前記機械翻訳装置は、入力装置から入力された第1言語の原文からキーワードを推定するキーワード推定部と、入力された原文全体に対して、前記キーワード推定部で推定されたキーワードのうち原文において最も文頭に近い位置に存在する第1言語の語が翻訳後の第2言語の訳文の文頭に最も近い位置に存在する訳文を生成する規則を選択して翻訳を行う翻訳部と、翻訳対象の原文や前記翻訳部による翻訳後の第2言語の訳文

を前記出力装置に出力処理する出力処理部とを備える機械翻訳装置。

【請求項4】

機械翻訳プログラム、翻訳に必要な知識情報・規則を蓄積した翻訳辞書部を記憶した記憶装置と、前記機械翻訳プログラムを演算実行する演算制御装置とを備えたコンピュータに用いられる機械翻訳プログラムにおいて、前記コンピュータに、入力装置から入力された第1言語の原文に対してユーザから指定された第1言語の語句を受け付ける手順と、入力された原文全体に対して、指定された第1言語の語句またはその語句を含むフレーズの訳語が翻訳後の第2言語の訳文の文頭に最も近い位置に存在する訳文を生成する規則を選択して翻訳を行う手順と、翻訳対象の原文や前記翻訳部による翻訳後の第2言語の訳文を出力装置に出力処理する手順とを実行させるための機械翻訳プログラム。

10

【請求項5】

機械翻訳プログラム、翻訳に必要な知識情報・規則を蓄積した翻訳辞書部を記憶した記憶装置と、前記機械翻訳プログラムを演算実行する演算制御装置とを備えたコンピュータに用いられる機械翻訳プログラムにおいて、前記コンピュータに、入力装置から入力された第1言語の原文からキーワードを推定する手順と、入力された原文全体に対して、推定されたキーワードのうち原文において最も文頭に近い位置に存在する第1言語の語が翻訳後の第2言語の訳文の文頭に最も近い位置に存在する訳文を生成する規則を選択して翻訳を行う手順と、翻訳対象の原文や前記翻訳部による翻訳後の第2言語の訳文を前記出力装置に出力処理する手順とを実行させるための機械翻訳プログラム。

20

【請求項6】

機械翻訳プログラム、翻訳に必要な知識情報・規則を蓄積した翻訳辞書部を記憶した記憶装置と、前記機械翻訳プログラムを演算実行する演算制御装置とを備え、機械翻訳を行う機械翻訳方法において、入力装置から入力された第1言語の原文に対してユーザから指定された第1言語の語句を受け付けるステップと、入力された原文全体に対して、指定された第1言語の語句またはその語句を含むフレーズの訳語が、翻訳後の第2言語の訳文の文頭に最も近い位置に存在する訳文を生成する規則を選択して翻訳を行うステップと、翻訳対象の原文や前記翻訳部による翻訳後の第2言語の訳文を出力装置に出力処理するステップとを備える機械翻訳方法。

【請求項7】

機械翻訳プログラム、翻訳に必要な知識情報・規則を蓄積した翻訳辞書部を記憶した記憶装置と、前記機械翻訳プログラムを演算実行する演算制御装置とを備え、機械翻訳を行う機械翻訳方法において、入力装置から入力された第1言語の原文からキーワードを推定するステップと、入力された原文全体に対して、キーワードのうち原文において最も文頭に近い位置に存在する第1言語の語が翻訳後の第2言語の訳文の文頭に最も近い位置に存在する訳文を生成する規則を選択して翻訳を行うステップと、翻訳対象の原文や前記翻訳部による翻訳後の第2言語の訳文を前記出力装置に出力処理するステップとを備える機械翻訳方法。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、第1言語の原文を第2言語の訳文に翻訳を行う機械翻訳装置、プログラム及び方法に関する。

40

【背景技術】

【0002】

コンピュータを利用して、第1言語（原言語）の文章を第2言語（目的言語）の文章に翻訳する機械翻訳装置においては、翻訳対象文書である原文文書の意図を的確に反映した訳文を生成することが非常に重要である。このような目的に対して、従来、以下の3つの方法が提案されてきた。

【0003】

第1の方法は、原文文書の分野を自動推定し、その分野に適した訳語を翻訳時に使用す

50

るというものである。複数の分野にまたがる文書に関しても分野情報を利用して言語解析を行うようにしたものとして、原文文書から使用頻度の高い語及び関連性の高い語(共起関係にある語)を統計的手法を用いて抽出し、語と分野とをあらかじめ関連付けた対応表を使用して、原文文書の分野を推定するようにしたものがある(例えば、特許文献1参照)。これによって、原文文書中の特定の語に対して複数の訳語が存在する場合でも、原文文書の分野に適した訳語が選択され、原文の意味が訳文に正確に反映される。

【0004】

第2の方法は、原文文書の丁寧度などの雰囲気解析し、その雰囲気にあった表現方法を用いて訳文を作成するというものである。このようなものとしては、入力文の内容を分析して得られる対話の親密度に応じて、訳語や敬語を調整・変更するようにしたものがある(例えば、特許文献2参照)。例えば、原文文書で用いられる表現を解析して、解析結果から丁寧さよりも親密さを重視した表現の頻度が高いことが判明した場合は、訳文で使用する表現にも親密さを重視した、より口語的な表現を用いる。これによって、メール・チャットなどでは口語表現を採用し、論文・レポートなどでは硬い表現を使用するなど、原文の雰囲気を反映した訳文を作成することができる。

10

【0005】

第3の方法は、翻訳を行った結果である訳文において、目的言語として不自然な構文が用いられている場合、より自然な訳文が生成できるように原文の構文を変換するというものである。このようなものとして、英日翻訳を対象とし、訳文である日本語文で無生物が主語になっている場合、変換のための翻訳規則に基づいて、原文である英文の構文を無生物を主語にしない構文に変換するようにしたものがある(例えば、特許文献3参照)。例えば、"A five-minute walk will take me to the station."という英文は、英文の構文に忠実に翻訳すると「5分歩行は駅へ私を連れていく」という、無生物が主語となった不自然な日本語になってしまう。この英文を"If I have a five-minute walk, I will get to the station."と変換してから日本語に翻訳すると、「5分歩けば私は駅に着く」という日本語として自然な文が得られる。このような変換のための翻訳規則は、辞書に登録された動詞に事前に登録されているとしている。このようにして変換した原文を翻訳することで、目的言語としてより自然な構文の訳文を作成することができる。

20

【0006】

このように、翻訳対象文書である原文文書の意図を的確に反映した訳文を生成するという目的に対して、従来の技術では、正確で文脈に即した用語の使用、文章の雰囲気にあった用語の選択、特に英日翻訳の訳文での無生物主語の回避が試みられてきた。

30

【特許文献1】特開平05-135094号公報

【特許文献2】特開2003-288338号公報

【特許文献3】特開平5-40782号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかしながら、従来の技術が対象としていたのは、単語及び1文単位での文にとどまっているので、原文文書を構成する各文の関連性や段落及び文書全体レベルでの論理の流れなどを訳文に反映することはできない。すなわち、前述した従来の第1の方法によって原文の分野に適した訳語を使用し、従来の第2の方法によって原文の雰囲気を反映した言い回しを採用したとしても、これらの方法はあくまでも1文単位での表現形態に注目したものであるため、各文をつなぐ論理展開には注目していない。その結果、1文単位では正しい訳文になっていながら、全体として何を言いたいのかが非常に分かりにくい訳文文書が生成されてしまうという問題が発生することがある。

40

【0008】

この問題は、原文文書が日本語である場合に特に顕著である。この理由について、日英翻訳の場合を例にして説明する。日本語として自然な表現を忠実に英文に翻訳すると、英文で多用される論理展開形式に則った英文にはならない場合がある。英文として分かりや

50

すい論理展開は、多くの場合、段落のトピックワードとなるキーワードを主語としたトピックセンテンスを段落の最初に述べ、次の文において、トピックセンテンスの中で導入された概念を表す語、あるいはその関連語を主語として詳述していくスタイルをとる。副詞及び接続詞が文頭に来自る場合や、発言を引用する場合などを除いて、主語は多くの場合文頭に位置する。従って、1つの段落内での各文の文頭の語を辿っていけば、文章の論理展開上重要な語であるキーワードのつながりが分かり、論理の展開が分かるようになっている。

【0009】

一方、日本語として自然な表現を従来の方法で英文に翻訳した翻訳結果においては、上述した英文の自然な論理展開になっているとは限らない。これについて、以下説明する。いま、段落のトピックワードを「キーワード1」とし、次の文でキーワード1が原因となっている結果(キーワード2)が生じるという内容を記述する場合を例に挙げて説明する。この内容を英文で記述する場合は、「KEYWORD1 causes KEYWORD2.」というように、多くの場合、段落のトピックワードであるKEYWORD1を主語とする構文を採用し、その結果、KEYWORD1は文頭に位置する。一方、この内容を日本語で表した文では、「キーワード1によってキーワード2が発生した」、「キーワード1を行ったので、キーワード2が発生した」というようになる。これらの文では、キーワード1が文頭に位置するのは英語の場合と同じだが、因果関係が副詞句、条件節などを用いて表されている。このような傾向を持つ日本語文を、日本語の構文をそのまま活かして英文に翻訳すると、翻訳結果の英文では「KEYWORD2 is caused by KEYWORD1.」、「KEYWORD2 occurred because KEYWORD1 was carried out.」というように、キーワード2の「結果」が主語となって文頭に位置し、キーワード1の「原因」は、副詞句、条件節の中に埋もれてしまう。その結果、KEYWORD1がトピックワードであることが分かりにくくなり、論理展開がわかりにくい英文となってしまう。

【0010】

このような問題に対して、従来の技術による第3の方法を翻訳方向を変えて単純に適用したとしても、問題を解決することはできない。従来の技術による第3の方法は英日翻訳を対象としており、訳文である日本語文で無生物が主語とならないように、原文の構文を変換している。そこで、この考え方を日英翻訳に応用すると、無生物が主語である英文が生成できるように日本語原文を変換することになる。しかし、何を主語にすべきかを機械的に決めることはできない。

【0011】

前述の例で言うと、英文として分かりやすい論理展開の文を作成するために、キーワード1とキーワード2とのどちらを主語、すなわち文書のトピックワードとすべきかに関して、翻訳対象の1文のみからは明確な判断基準が得られない。この判断を行うには、段落あるいは文書全体に関わる情報が別途必要となるからである。従って、論理展開がわかりやすい英文を得るという目的のために、従来の技術による第3の方法を翻訳方向を変えて単純に適用することはできない。

【0012】

本発明の目的は、第1言語の原文の論理展開が第2言語の訳文として自然なスタイルで明確に表現された訳文を作成することができる機械翻訳装置、プログラム及び方法を提供することである。

【課題を解決するための手段】

【0013】

本発明は、翻訳に必要な知識情報・規則を蓄積した翻訳辞書部を記憶した記憶装置を備えた機械翻訳装置において、前記機械翻訳装置は、入力装置から入力された第1言語の原文に対してユーザから指定された第1言語の語句を受け付ける主要キーワード指定部と、入力された原文全体に対して、前記主要キーワード指定部で指定された第1言語の語句またはその語句を含むフレーズの訳語が、翻訳後の第2言語の訳文の文頭に最も近い位置に存在する訳文を生成する規則を選択して翻訳を行う翻訳部と、翻訳対象の原文や前記翻訳

部による翻訳後の第 2 言語の訳文を出力装置に出力処理する出力処理部とを備える。

【発明の効果】

【0014】

本発明によれば、入力された第 1 言語の原文につき入力装置から論理展開上重要な語を指定すると、翻訳部は、指定された語またはその語を含むフレーズが文頭に最も近い位置に存在するように第 2 言語の訳文を生成するので、原文の論理展開上重要な語句が訳文においても論理展開上重要な位置に置かれる。このため、訳文の論理の展開が明確な訳文を作成することができ、原文の論理の展開が訳文言語として自然なスタイルで明確に表現された訳文となる。

【発明を実施するための最良の形態】

10

【0015】

以下、図面を参照しながら本発明の実施の形態を説明する。図 1 は本発明の実施の形態に係る機械翻訳装置のハードウェア構成を示すブロック構成図である。機械翻訳装置 11 は、例えば一般的なコンピュータに機械翻訳プログラムなどのソフトウェアプログラムがインストールされ、そのソフトウェアプログラムが演算制御装置 12 のプロセッサ 13 において実行されることにより実現される。

【0016】

演算制御装置 12 は機械翻訳に関する各種演算を行うものであり、演算制御装置 12 はプロセッサ 13 とメモリ 14 とを有し、メモリ 14 には翻訳に関する機械翻訳プログラム 15 が記憶され、プロセッサ 13 により処理が実行される際には作業エリア 16 が用いられる。演算制御装置 12 の演算結果等は出力装置 17 である表示装置 18 に表示出力され、また、通信制御装置 19 を介して通信ネットワークに出力される。

20

【0017】

入力装置 20 は演算制御装置 12 に情報を入力するものであり、例えば、マウス 21、キーボード 22、ディスクドライブ 23、通信制御装置 19 から構成され、例えば、マウス 21 やキーボード 22 は表示装置 18 を介して演算制御装置 12 に各種指令を入力し、キーボード 22、ディスクドライブ 23、通信制御装置 19 は翻訳対象の文書を入力する。

【0018】

すなわち、ディスクドライブ 23 は翻訳対象の文書のファイルを記憶媒体に入出力するものであり、通信制御装置 19 は機械翻訳装置 11 をインターネットや LAN などの通信ネットワークに接続するものである。通信制御装置 19 は LAN カードやモデムなどの装置であり、通信制御装置 19 を介して通信ネットワークと送受信したデータは入力信号又は出力信号として演算制御装置 12 に送受信される。さらに、演算制御装置 12 の演算結果や翻訳に必要な知識情報・規則を蓄積した翻訳辞書等を記憶するハードディスクドライブ (HDD) 24 が設けられている。

30

【0019】

図 2 は本発明の実施の形態に係る機械翻訳装置 11 の機能ブロック図である。図 2 に示す演算制御装置 12 内の各機能ブロックは、上述の機械翻訳プログラム 15 を構成する各プログラムに対応する。すなわち、プロセッサ 13 が機械翻訳プログラム 15 を構成する各プログラムを実行することで、演算制御装置 12 は、各機能ブロックとして機能することとなる。

40

【0020】

図 2 において、入力処理部 25 は、入力装置 20 から翻訳対象となる文書（以下、「原文文書」という）のデータ入力を受け付けるものである。例えば、原文文書は、ユーザ自身によりキーボード 22 等から直接入力されたり、記憶媒体を介してディスクドライブ 23 から入力される。なお、原文文書の入力方式については、これらに限定されるものではなく、任意の入力方式が選択可能である。

【0021】

文分割部 26 は、入力処理部 25 で入力処理された原文文書を文単位に分割し作業エリ

50

ア 16 に格納するものである。1 文単位で分割された文（以下、「原文」という）は、後述する制御部 27 に受け渡され、同じく後述する翻訳部 28 に送られて翻訳処理を受ける。

【0022】

制御部 27 は、後述する翻訳部 28、キーワード推定部 29、主要キーワード指定部 30 を制御するものである。これらを制御することで、入力された翻訳対象文の翻訳を行うと共に、翻訳対象文のキーワード推定、ユーザが指定したキーワードの条件に基づく翻訳文生成の制御を行う。キーワード推定部 29 は、原文文書のキーワードの自動推定を行うものであり、文書構造解析部 29a、キーワード推定処理部 29b から構成されている。ここでいうキーワードは、原文の論理展開に重要な役割を果たす語のことである。

10

【0023】

翻訳部 28 は、入力処理部 25 で入力処理され文分割部 26 によって 1 文単位に分割された原文の翻訳を、記憶装置 31 に記憶された翻訳辞書部 32 の基本辞書部 33 や用例辞書部 34 を参照しながら行うものである。翻訳辞書部 32 の基本辞書部 33 には、入力原文を第 2 言語の文に翻訳する際に必要な知識情報が格納されている。用例辞書部 34 には原文と訳文の対訳をペアで登録した用例が格納されている。これらの知識情報は、翻訳部 28 において翻訳時に参照される。

【0024】

翻訳辞書部 32 は、翻訳部 28 の翻訳処理に必要な知識情報・規則を記憶しており、例えば、翻訳部 28 が日英・英日双方向の翻訳処理を行うものとする、翻訳辞書部 32 は夫々の翻訳方向毎に、基本辞書部 33、用例辞書部 34 を有している。基本辞書部 33 は、語彙部 33a、形態素解析規則 33b、構文・意味解析規則 33c、変換規則 33d、構文生成規則 33e、形態素生成規則 33f からなる。基本辞書部 33 の構成は以下の通りである。

20

【0025】

(1) 基本辞書部

(a) 語彙部

少なくとも第 1 言語の語彙の各々についての活用情報、意味情報、分野情報、訳語情報、訳語毎の分野情報。

(b) 形態素解析規則

30

第 1 言語の入力文を形態素解析するための知識。

(c) 構文・意味解析規則

第 1 言語の入力文を形態素解析した後、構文的・意味的な解析を行うための知識。

(d) 変換規則

構文・意味解析された結果の第 1 言語の意味構造を第 2 言語の意味構造へ変換するための知識。

(e) 構文生成規則

第 2 言語の意味構造から第 2 言語の単語列を生成するための知識。

(f) 形態素生成規則

40

第 2 言語の語の活用を反映し、最終的な訳文を出力するための知識。

【0026】

主要キーワード指定部 30 は、ユーザが原文において入力装置 20 から入力処理部 25 を介して「主要キーワード」を指定する際に、その指定を受け付けるものである。ここでいう「主要キーワード」とは、原文の論理展開に重要な役割を果たし、かつ翻訳結果である訳文において、接続詞、修飾語、“He said that” などの発言を表す文字列などを除いて文頭に位置するべき語（原文及び訳文の双方）を指す。「主要キーワード」は、例えば入力装置 20 のマウス 21 等によるドラッグ操作によって指定される。

【0027】

出力処理部 35 は、翻訳部 28 によって生成された翻訳の結果を出力装置 17 に出力処

50

理するものである。例えば、出力装置 17 の表示装置 18 に表示することで翻訳結果を出力処理する。しかし、これに限定されるものではなく、表示装置 18 による出力に代え、あるいは、表示装置 18 による出力に加えて音声により翻訳結果を出力処理するようにしても良い。

【0028】

次に、本発明の実施の形態に係わる機械翻訳装置 11 における翻訳処理について説明する。図 3 は、本発明の実施の形態に係わる機械翻訳装置 11 の翻訳処理のフロー図である。本発明の実施の形態に係わる機械翻訳装置 11 を用いて翻訳処理を行う場合、まず、ユーザから入力装置 20 により入力された原文文書は、入力処理部 25 から入力され受け付けられる（ステップ S11）。入力された原文文書は文分割部 26 により、原文文書が 1 文単位に分割される（ステップ S12）。分割された各原文は、1 文ごとに翻訳部 28 に送られて一旦翻訳処理を受ける（ステップ S13）。

10

【0029】

翻訳部 28 では、翻訳辞書部 32 の各種辞書を参照し翻訳処理を行う。翻訳処理の詳細については、原文の形態素解析処理、原文の構文・意味解析、原文言語から訳文言語への言語変換処理、原文の構文から訳文の構文への構文生成処理、訳文としての形態素生成処理を受けて、最終的な翻訳結果である英文が生成される、というステップを踏む。

【0030】

すなわち、翻訳部 28 は語彙部 33a 及び形態素解析規則 33b を参照して、原文を言語で意味を持つ最小単位である形態素の単位に分割する。次に、構文・意味解析規則 33c を参照して、形態素に分割された原文の語の見出し語、訳語それぞれを構成する各語の構文・意味的な関係を解析する。構文・意味的な関係とは、例えば動詞とその目的語、前置詞とその目的語、意味上の主語と述語、修飾語と非修飾語などの関係を意味する。最後に、変換規則 33d、構文生成規則 33e 及び形態素生成規則部 33f を参照して対応する英文を生成する。翻訳部 28 での翻訳処理は、入力されたすべての原文について処理が終了するまで繰り返し実行される（ステップ S14）。

20

【0031】

なお、以上の説明では、翻訳処理の一形態であるルール（翻訳規則）ベースの翻訳処理の内容について説明したが、本発明による翻訳形態は、ルールベースの翻訳処理に限定されるものではなく、例えば用例ベースの翻訳処理を採用することも可能である。なお、用例ベースの翻訳においては、原文と訳文の対訳をペアで登録した用例辞書部 34 を用いて、翻訳対象の文と類似度の高い原文とを用例辞書部 34 の中から検出し、その原文とペアになっている訳文を翻訳結果として出力する、という翻訳処理を行う。

30

【0032】

入力されたすべての原文に対して翻訳処理が終了すると、制御部 27 は翻訳結果を出力処理部 35 を介して表示装置 18 に表示する（S15）。ユーザは表示装置 18 に表示された翻訳結果である訳文を検討し、翻訳結果が満足であるかどうかを判断する（S16）。翻訳結果に満足ならば処理は終了となり、満足でないならば、キーワード推定部 29 での処理に進む。

【0033】

キーワード推定部 29 では、原文文書のキーワード推定が行われる（ステップ S17）。キーワード推定の方法には、例えば特開平 8-44763 号公報で開示されているような既存の技術を利用することができる。キーワード推定に際しては、単語の出現頻度、共起情報などの統計情報、タグ付き文書においてはタグ情報、及び書式付文書においては書式情報などの文書構造情報、文章・段落中にて最初や最後に位置するなどの位置情報、接続詞などの文どうしの論理関係をあらわす文間情報、翻訳部 28 での処理結果から得られる各文の構文・意味情報などの情報を用いる。これらの情報から、原文において論理展開上重要な語であるキーワードを推定する。なお、ここで説明したキーワード推定の手段はあくまでも一例であり、原文文書のキーワード推定をできるものであればいずれの手法を用いてもよい。

40

50

【0034】

続いて、主要キーワード指定部30において、キーワード推定部29で推定されたキーワードをユーザは参考にして、ユーザが入力装置20より「主要キーワード」の指定を行う（ステップS18）。ここで指定する「主要キーワード」とは、前述したように、原文の論理展開に重要な役割を果たし、かつ翻訳結果である訳文において、接続詞、修飾語、“He said that”などの発言を表す文字列などを除いて文頭に位置するべき、原文及び訳文の双方の語である。続いて翻訳部28において、前述した「主要キーワード」の条件を満たす英文の生成が試みられる（ステップS19）。

【0035】

すなわち、翻訳部28での翻訳辞書部32における基本辞書部33の語彙部33aおよび変換規則33dを用いた言語変換処理及び構文生成規則33eを用いた構文生成処理において、言語変換及び英文構文生成の方法が複数種類ある場合には、翻訳部28は、指定された「主要キーワード」が、できるだけ文頭に近い位置に存在するような方法を選択する。これら複数種類の処理方法は、基本辞書部33に格納された各知識情報および種翻訳規則の組み合わせによって発生する。これらの規則としては、語彙ごとに特有な知識情報および変換規則が語彙部33aに、語彙には依存しない一般的な変換規則が変換規則33dに、英文として正しい構文を生成するための規則が構文生成規則33eに格納されている。これらの規則の中から、翻訳部28によって主要キーワードが可能な限り文頭に近い位置に存在する条件に合った訳文を生成する翻訳規則が選択され、翻訳時にこれらの規則が優先的に用いられる。

【0036】

一方、翻訳処理において用例ベースの翻訳処理を用いる場合には、翻訳部28は、用例辞書部34を参照し、訳文候補として挙げた英文用例の中から、主要キーワードが可能な限り文頭に近い位置に存在する条件に最も合致する英文用例の選択を行う。

【0037】

制御部27は主要キーワードが可能な限り文頭に近い位置に存在する条件に合った翻訳結果を出力処理部35を介して表示装置18に表示する（S20）。ユーザは表示装置18に表示された翻訳結果である英文を検討し、翻訳結果が満足であるかどうかを判断する（S21）。翻訳結果に満足ならば処理は終了となり、満足でないならば主要キーワードを代えるかどうかを判断し（S22）、主要キーワードを代える場合はステップS18に戻り新たな主要キーワードを指定する。一方、主要キーワードを代えない場合は、規格翻訳による翻訳をあきらめ処理を終了する。

【0038】

以下、図3に示す翻訳処理内容について具体例を参照して詳述する。いま、以下に示す3つの原文S1、S2、S3に対して、翻訳部28において翻訳処理を行うものとする。なお、以下の説明では日英翻訳を一例にして説明するが、原文及び訳文言語はこれに限らず、言語特性の違いによって、原文の構文を活かした翻訳を行うと、訳文文書にて原文の論理展開が分かりにくくなるすべての言語に対して、本発明を適用することができる。

【0039】

(原文)

S1：磁気ベアリングは航空宇宙産業向けの用途に開発されてきたが、ごく最近、エネルギー貯蔵システムの心臓部としての実現性が実証された。

S2：このようなブレイクスルーが見られた原因は、より強力な永久磁石が最近開発されたことにある。

S3：そのような磁石が10ポンドあれば、2トンのローターを支持できる。

【0040】

これらの原文の論理展開について説明する。これらの原文に対して、従来の方法で翻訳を行うと、翻訳結果はT1-a、T2-a、T3-aとなる。

【0041】

(従来の方法による翻訳結果)

T1-a. Although the magnetic bearing had been developed for aerospace and aircraft works, the implementability as a core of an energy storage system was proved very much recently.

T2-a. There is a cause by which such a breakthrough was seen in the more powerful permanent magnet having been developed recently.

T3-a. If there are 10 pounds of such magnets, a 2t rotor can be supported.

原文中の下線を引いた語は、論理展開上重要な語であり、かつ文頭に位置する語である。これらの語を辿ってみると、原文の論理展開を辿ることができる。一方、原文にて下線を引いた語の翻訳結果における文中の位置を見てみると、原文では各文の文頭にあった語が、訳文T2、T3では文中に埋もれてしまっている。その結果、訳文の英文は、1文単位で意味は通じるが、全体の論理の流れが英文として不明確になってしまっている。

【0042】

従来の方法を用いると、翻訳結果にこのような問題があった、これに対し本発明では、以下で説明する処理を行って、論理展開が明確な英文を作成する。以下に、キーワード推定部29により推定されたキーワードが表示された状態の原文を示す。推定されたキーワードを「」付の下線で示している。ここでいうキーワードとは、名詞、動詞などの自立語から構成される、一般的な意味でのキーワードである。

【0043】

(原文：推定キーワードが表示された状態)

S1：「磁気ベアリング」は「航空宇宙産業」向けの用途に「開発」されてきたが、ごく最近、「エネルギー貯蔵システム」の「心臓部」としての「実現性」が実証された。

S2：このような「ブレイクスルー」が見られた原因は、より強力な「永久磁石」が最近「開発」されたことにある。

S3：そのような「磁石」が10ポンドあれば、2トンの「ローター」を「支持」できる。

【0044】

次に、「」付の下線の推定されたキーワードを参考にして、ユーザは原文の論理展開に重要な役割を果たし、かつ翻訳結果である訳文において、接続詞、修飾語、“He said that”などの発言を表す文字列などを除いて文頭に位置すべき語である「主要キーワード」を指定する。この一例では「ブレイクスルー」と「磁石」を指定したとする。以下に、原文S2、S3で「主要キーワード」が指定された状態を示す。指定された「主要キーワード」を[]付の下線で示す。

【0045】

(原文：「主要キーワード」を指定した状態)

S2：このような[ブレイクスルー]が見られた原因は、より強力な永久磁石が最近開発されたことにある。

S3：そのような[磁石]が10ポンドあれば、2トンのローターを支持できる。

【0046】

この一例では、「このような」「そのような」というような修飾句は除いて指定したが、このような修飾句を一緒に指定してもかまわない。

【0047】

このように「主要キーワード」を指定した原文に対して、指定した「主要キーワード」が最も文頭に近い位置に位置するような翻訳規則を翻訳辞書部32の変換規則33d及び構文生成規則33eから選択して、優先的に使用して翻訳を行った結果を以下に示す。具体的には、以下のT2-bでは、「(がみられた)原因は」という日本語を“The reason for ~”という英語ではなく、“~ is due to”という英語に変換する規則が変換規則33dから選択され、T3-bでは、「~があれば」という日本語を“If there are ~”という英語ではなく、“~ could”という英語構文を生成する規則を構文生成規則33eから選択している。これによって以下の各文では、指定した主要キーワードは文頭に近い位置に存在し、英文として理解しやすい論理展開となっている。

【0048】

(本発明による翻訳結果：「主要キーワード」指定後)

T2-b. [The breakthrough] is due to the recent development of stronger permanent magnets.

T3-b. Only ten pounds of [such magnets] could support two tons of rotor.

以上まとめると、本発明による機械翻訳装置においては、ユーザが原文において指定した語またはその語を含むフレーズ（ここでは「ブレイクスルー」）ができるだけ文頭に近い位置に存在するような訳文の生成方法を選択する。例えば、以下の表1に挙げたような翻訳結果候補1～4があった場合、本発明による機械翻訳装置では、ユーザが原文において指定した語（ブレイクスルー）が最も文頭に近い位置に存在する訳文（翻訳結果候補2）が翻訳結果となるように訳文の生成方法を選択する。

10

【表1】

原文(S2)	このようなブレイクスルーが見られた原因は、より強力な永久磁石が最近開発されたことにある。	網掛けの語の文頭からの位置	本発明で訳文に採用される順番
翻訳結果候補1 (T2-a)	<u>There is a cause</u> by which <u>such a breakthrough</u> was seen in the more powerful permanent magnet having been developed recently.	5番目	4
翻訳結果候補2 (T2-b)	<u>The breakthrough</u> is due to the recent development of stronger permanent magnets.	1番目	1
翻訳結果候補3	<u>The reason why</u> <u>such a breakthrough</u> was seen is that more powerful permanent magnets were developed recently.	3番目	3
翻訳結果候補4	<u>The cause of</u> <u>such a breakthrough</u> depends on the recent development of stronger permanent magnets.	3番目	2

20

30

【0049】

なお、ユーザが原文において指定した語が文頭に近い位置に存在する訳文である翻訳結果候補2が何らかの理由で生成できない場合には、翻訳部28は、指定した語を含むフレーズが最も文頭に近い訳文を選択する。表1で示した一例では、翻訳結果候補3、4がこれに相当する。翻訳結果候補3、4では、原文において指定した語（ブレイクスルー）が文頭から同じ位置に存在する。この場合は、原文において指定した語を含むフレーズが主語となっている場合に、翻訳部28は、主語の語数が少ない翻訳結果候補を選択する。ここでは、訳文結果候補3の主語は“The reason why such a breakthrough was seen”であり、訳文結果候補4の主語は“The cause of such a breakthrough”であり、訳文結果候補4の主語の語数が訳文結果候補3の主語の語数より少ないので、翻訳結果候補4を生成することになる。

40

【0050】

以上の説明は、原文を一旦翻訳した後にキーワード推定部29でキーワードの推定を行い、ユーザが入力装置20で推定されたキーワードの中から「主要キーワード」を指定して再度翻訳を行うようにしたが、最初からキーワードの自動推定を行い、最も文頭に近いキーワードをデフォルトで「主要キーワード」とし、訳文を生成するような翻訳モードを翻訳部28に設けることもできる。

50

【0051】

このような翻訳モードによって、必ずしも論理展開が明確な訳文が作成できるとは限らないが、情報は既知のものから未知のものという順で提示される場合が多いことを考慮すると、文頭に近いキーワードが前の文の内容を受けた「主要キーワード」である可能性は比較的高いと考えられるため、デフォルトモードとしての価値は高いと考えられる。もちろん、デフォルトの「主要キーワード」をユーザが指定し直して再度翻訳することもできる。また、キーワード推定部29を省略し、原文を入力した後に、ユーザがその原文に対して、入力装置20から「主要キーワード」語句を指定し翻訳するようにしてもよい。この場合は、図3のステップS11の後に、主要キーワードの語句を指定するステップを設け、ステップS16～ステップS20を省略することになる。

10

【0052】

次に、原文S2、訳文T2-a、T2-bを一例に用いて、翻訳部28で行われる各処理の詳細について説明する。なお、ここで行っている処理はルールベースの翻訳処理の一例である。

【0053】

図4は、本発明の実施の形態に係わる機械翻訳装置の翻訳処理（図3のステップS13）における形態素解析・構文解析の結果として得られる木構造を示す説明図である。入力された原文は、原文S2の「このようなブレイクスルーが見られた原因は、より強力な永久磁石が最近開発されたことにある。」である。図4の木構造では、文を構成する自立語（ノード）と、各自立語間の構文及び意味上の関係（リンク）が明示されている。なお、簡潔のため、キーワード以外の語に関するリンクを適宜省略した形で書いている。なお、強調表示してあるノードとリンクは、後述の図6を説明する際に用いる。

20

【0054】

図5は、図4に示した原文の構文及び意味上の関係をできるだけ忠実に反映して、翻訳処理（図3のステップS13）における言語変換処理及び構文生成処理を行った結果得られた木構造を示す説明図である。この構文では、「breakthrough」は「see」の目的語となっており、日本語の構文がほぼそのまま英文の構文に変換されている。

【0055】

図6は、図4の原文において「ブレイクスルー」をキーワードとして指定して翻訳処理（図3のステップS19）における言語変換処理及び構文生成処理を行った結果得られた木構造を示す説明図である。ここで優先された言語変換及び構文生成の規則を図7を用いて説明する。図7（a）は原文の部分木構造で、図7（b）は規則適用後の部分木構造である。翻訳規則の概要は、「が見られた原因は～ことにある」という部分木がある場合、「見られた」の目的語である「ブレイクスルー」を主語とし、「～こと」であらわされる名詞節を名詞句に変換し、最終的に「breakthrough is due to 名詞句」という英語構文を生成する、というものである。翻訳規則は、具体的には、以下のようなテキスト形式で記述されている。この例は、「～が見られた原因は」という日本語を、「～ is due to」という英語に変換するための規則である。

30

【0056】

原因(syusyoku_*1(jyoshi_*2)=*2 is due to
[*1.prop=keisiki_v]

40

上記の*1と*2は単語を示す。「syusyoku_」は、「syusyoku_」の右側の文字列が「原因」を修飾していることを示し、「jyoshi」は、*1の単語と*2の単語が助詞（この例では「が」）でつながっていることを示している。[*1.prop=keisiki_v]の部分は、この規則が適用されるための条件で、:1の単語の属性が形式名詞(keisiki_V)であることを示している。形式動詞とは、「が見られる」「～と聞く」「～という」などのように、具体的な動作としての意味を失っている動詞のことである。例えば、keisiki_v：形式動詞（見る、聞く、言う・・・）である。

【0057】

以上説明したように、本発明の実施の形態では、ユーザが原文において指定した語また

50

はその語を含むフレーズができるだけ文頭に近い位置に存在するような訳文の生成方法を選択して原文を翻訳する。これによって、原文の論理展開上重要な語が訳文においても論理展開上重要な位置に置かれるため、論理の展開が明確な訳文を作成することができる。すなわち、原文の論理の展開が訳文言語として自然なスタイルで明確に表現された訳文を作成することができる。

【0058】

また、キーワード推定部29により原文の論理展開上重要な語の自動推定を行い、ユーザによる「主要キーワード」の指定の際に、推定された語句を参考にすることもできる。これによって、ユーザが語を指定する際の判断の負担を軽減することができる。

【0059】

また、キーワード推定部29により、原文の論理展開上重要な語の自動推定を行い、推定された語句のうち、最も文頭に近いキーワードを「主要キーワード」のデフォルト値とした訳文を生成するような翻訳モードを設けることもできる。これによって、「主要キーワード」の指定を行わない場合でも、訳文の作成に当たって原文の論理の展開に一応の配慮を行うことができる。

【図面の簡単な説明】

【0060】

【図1】本発明の実施の形態に係る機械翻訳装置のハードウェア構成を示すブロック構成図。

【図2】本発明の実施の形態に係る機械翻訳装置の機能ブロック図。

【図3】本発明の実施の形態に係る機械翻訳装置の翻訳処理のフロー図。

【図4】本発明の実施の形態に係る機械翻訳装置の翻訳処理（図3のステップS13）における形態素解析・構文解析の結果として得られる木構造を示す説明図。

【図5】図4に示した原文の構文及び意味上の関係をできるだけ忠実に反映して、翻訳処理（図3のステップS13）における言語変換処理及び構文生成処理を行った結果得られた木構造を示す説明図。

【図6】図4の原文において「ブレイクスルー」をキーワードとして指定して翻訳処理（図3のステップS19）における言語変換処理及び構文生成処理を行った結果得られた木構造を示す説明図。

【図7】図4の原文において「ブレイクスルー」をキーワードとして指定したときの優先された言語変換及び構文生成の規則の説明図。

【符号の説明】

【0061】

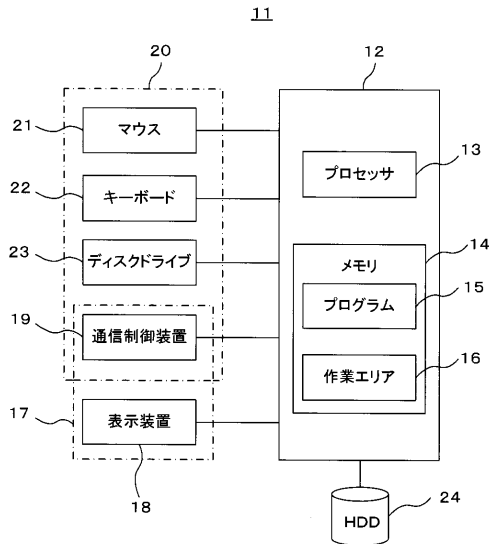
11…機械翻訳装置、12…演算制御装置、13…プロセッサ、14…メモリ、15…機械翻訳プログラム、16…作業エリア、17…出力装置、18…表示装置、19…通信制御装置、20…入力装置、21…マウス、22…キーボード、23…ディスクドライブ、24…ハードディスクドライブ、25…入力処理部、26…文分割部、27…制御部、28…翻訳部、29…キーワード推定部、30…主要キーワード指定部、31…記憶装置、32…翻訳辞書部、33…基本辞書部、34…用例辞書部、35…出力処理部

10

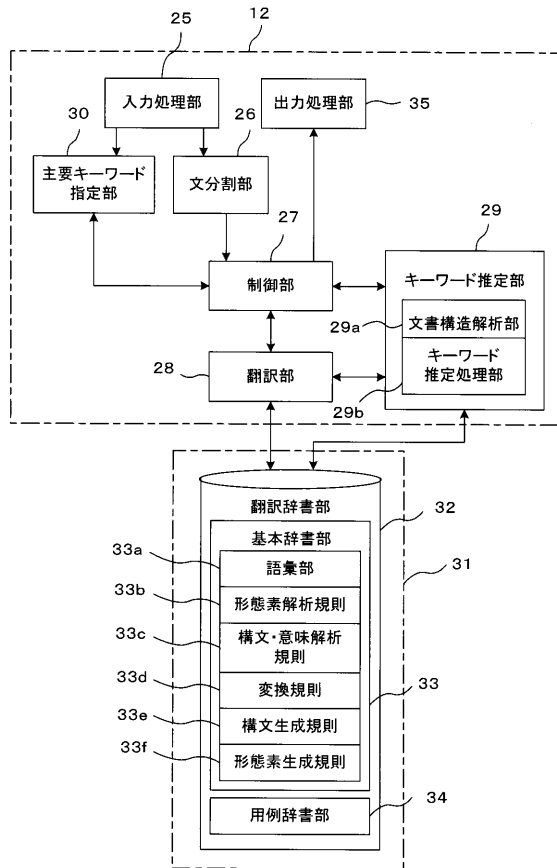
20

30

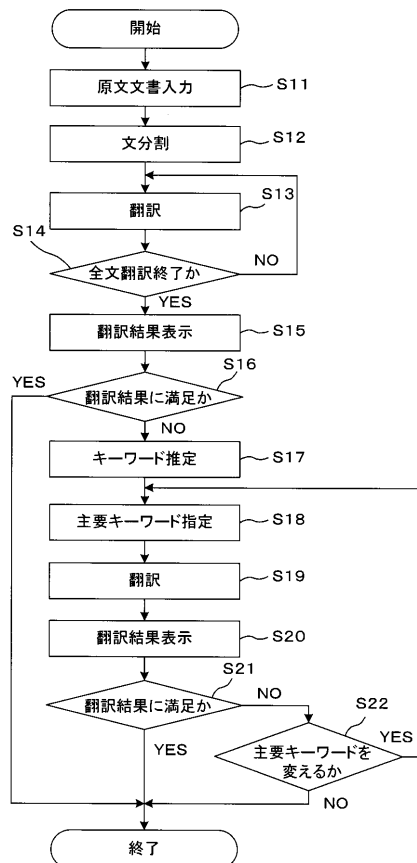
【図 1】



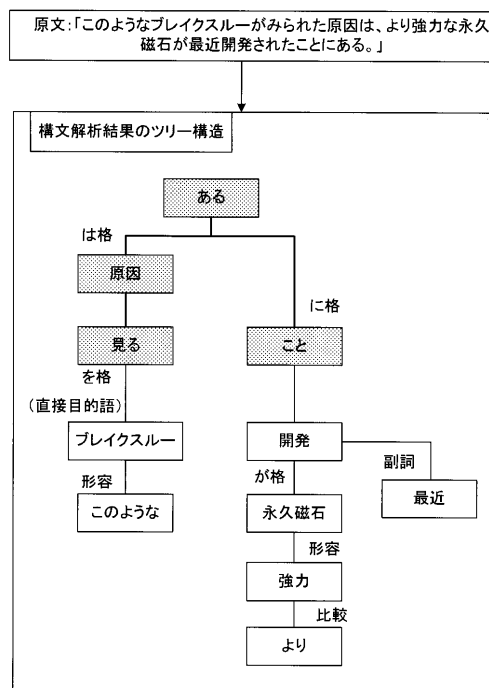
【図 2】



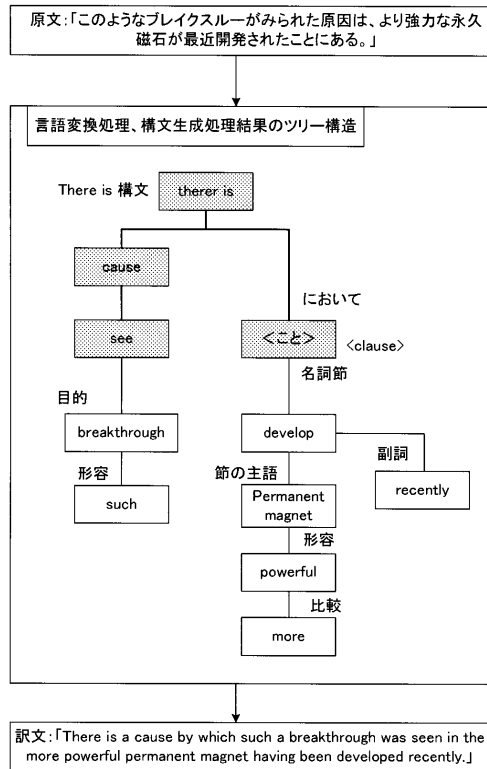
【図 3】



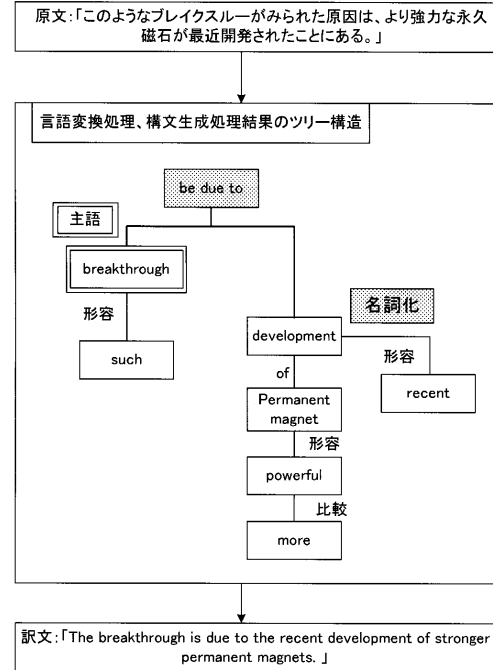
【図 4】



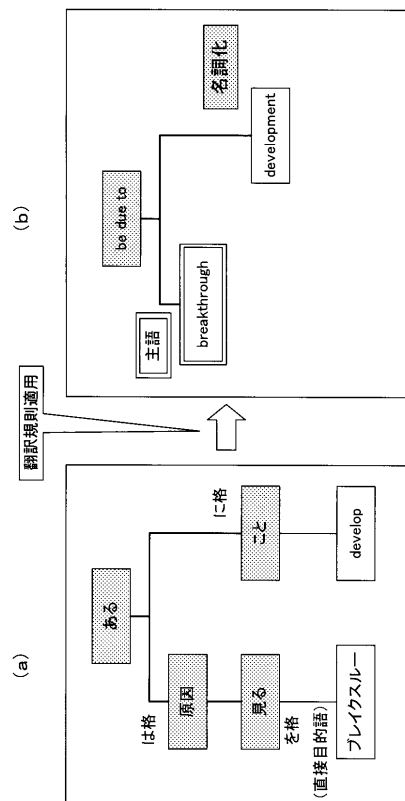
【図 5】



【図 6】



【図 7】



フロントページの続き

(56)参考文献 特開2002-328917 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/27-17/28