(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2018年5月31日(31.05.2018)

WIPO | PCT

(10) 国際公開番号
WO 2018/097022 A1

(51) 国際特許分類:
G06F 17/28 (2006.01)

(21) 国際出願番号： PCT/JP2017/041249

(22) 国際出願日： 2017年11月16日(16.11.2017)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(30) 優先権データ：
特願 2016-227583 2016年11月24日(24.11.2016) JP

(71) 出願人: 国立研究開発法人情報通信研究機構 (**NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY**) [JP/JP]; 〒1848795 東京都小金井市貫井北町４－２－１ Tokyo (JP).

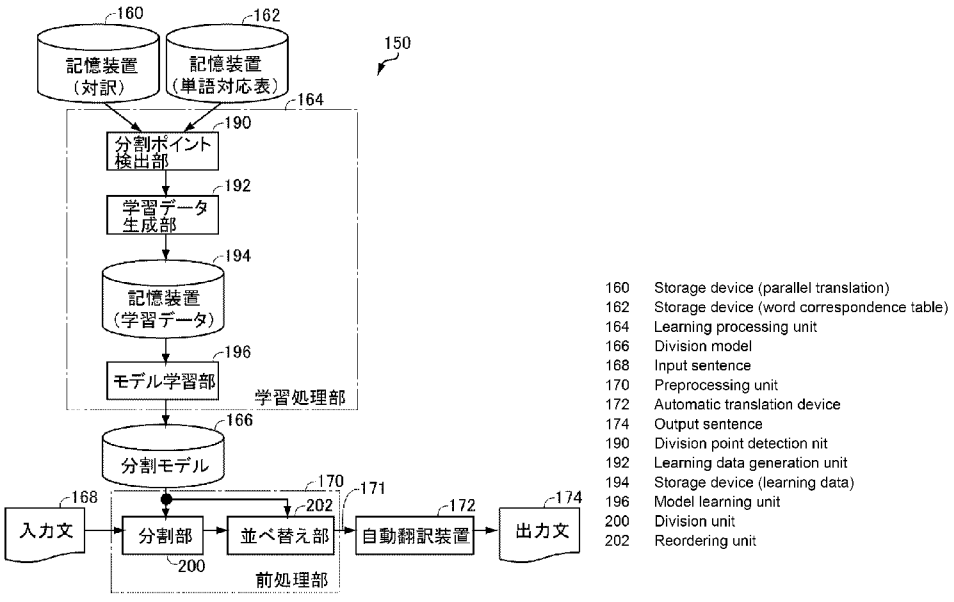
(72) 発明者: 富士 秀(**FUJI, Masaru**); 〒1848795 東京都小金井市貫井北町４－２－１ 国立研究開発法人情報通信研究機構内 Tokyo (JP). 内山 将夫(**UCHIYAMA, Masao**); 〒1848795 東京都小金井市貫井北町４－２－１ 国立研究開発法人情報通信研究機構内 Tokyo (JP).

(74) 代理人: 清水 敏(**SHIMIZU, Satoshi**); 〒5300002 大阪府大阪市北区曽根崎新地二丁目５番３号 堂島ＴＳＳビル４階 Osaka (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,

(54) **Title:** AUTOMATIC TRANSLATION PATTERN LEARNING DEVICE, AUTOMATIC TRANSLATION PREPROCESSING DEVICE, AND COMPUTER PROGRAM

(54) 発明の名称： 自動翻訳パターン学習装置、自動翻訳の前処理装置、及びコンピュータプログラム



(57) **Abstract:** [Problem] Provided is an automatic translation pattern learning device for learning patterns for, prior to automatic translation, transposing constituent component strings of an original text into an order suitable for automatic translation. [Solution] A learning processing unit 164 includes: a dividing point detection unit 190 for detecting in a parallel translation, as the division position of two sentences, the position at which the corresponding relationship between

WO 2018/097022 A1

[続葉有]

CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告(条約第21条(3))

the order of occurrence of constituent components of a sentence in the original language and the order of occurrence of constituent components of the corresponding translated sentence changes; a learning data generation unit 192 for generating, as learning data for mechanical learning, data comprising a constituent component string of a continuous prescribed number of items within the sentence in the original language, and correct answer information indicating whether the division position is within the constituent component string; and a model learning unit 196 for, when an input constituent component string of a prescribed number of units in a sentence in the original language is provided by machine learning, learning a division model 166 for determining whether the division position is at the prescribed position in the constituent component string, using the generated learning data.

(57) 要約:【課題】自動翻訳に先立って、原文の構成部品列を自動翻訳に適した順序に入れ替えるパターンを学習する自動翻訳パターン学習装置を提供する。【解決手段】学習処理部１６４は、対訳のうち、原言語の文の構造部品の出現順序と、対応する訳文の構造部品の出現順序との対応関係が変化する位置を両者の文の分割位置として検出する分割ポイント検出部１９０と、原言語の文内の、連続する所定個数の構造部品列と、その中に分割位置があるか否かを示す正解情報とからなるデータを機械学習の学習データとして生成する学習データ生成部１９２と、生成された学習データを用いて、機械学習により、原言語の文の所定個数の入力構造部品列が与えられたときに、その中の所定位置に分割位置があるか否かを判定する分割モデル１６６の学習を行うモデル学習部１９６を含む。

明　細　書

発明の名称 ：

自動翻訳パターン学習装置、自動翻訳の前処理装置、及びコンピュータプログラム

技術分野

[0001]　この発明は自動翻訳技術に関し、特に、定型性の高い文書に対する自動翻訳の精度を向上させる技術に関する。

背景技術

[0002]　自動翻訳では、定型的な文書を翻訳することが多い。典型的には特許出願書類のうちの権利範囲を定める部分の翻訳である。例えば、米国における先行技術調査を日本の技術者が行う場合、問題となるのはクレームの部分である。英語の文で記載された米国特許のクレームを技術者、弁護士、及び弁理士等がそのまま理解できれば理想的であるが、限られた時間に大量の英文クレームを読む負担は大きい。そのため、英文クレームを日本語に精度高く翻訳する技術が望まれている。

[0003]　特許文献１において、そうした技術が提案されている。特許文献１に記載された技術では、英文クレームと、その英文クレームに対応する日本語クレームが共通した定型性を持っていることを利用する。

[0004]　より具体的には、日英翻訳では、原文及び訳文の間で各構造部品の出現順序の対応関係が入替わることがしばしば起こる。例えば図１（Ａ）に示される対訳３０を考える。図１（Ｂ）に示すように、この対訳３０の構造部品列の対応関係を考えると、日本語の構造部品列４０、４２及び４４が、英語の構造部品列５４，５２及び５０にそれぞれ翻訳されており、構造部品列４０及び構造部品列４４の出現順序が、英語では構造部品列５０及び５４となって互いに入れ替わっている。図２（Ａ）に示される対訳７０についても同様である。日本語では構造部品列８０，８２及び８４がこの順序で出現しているのに対し、英語では構造部品列８０及び８４に対応する部分が入れ替わり

、構造部品列９０，９２及び９４の順序で出現している。

[0005] 特許文献１に開示された技術は、英文クレームと日本語クレームとのこのような関係を利用することで、自動翻訳の精度の向上を図っている。すなわち、予め英文クレームの構造部品と日本語クレームとの双方について、構造部品の並びのパターンを特定し、構造部品の対応関係（並べ替え）のパターンを特定する。このパターンにより、対応する英文クレームと日本語クレームとについて、構造部品の対応関係が特定できる。翻訳時に英文クレームが入力されると、英文クレームのパターンを特定し、そのパターンにしたがって英文クレームを構成要素に分割する。特定されたパターンに対応する日本語クレームのパターンを特定し、英文クレームの構成要素を日本語クレームのパターンに応じて並べ替える。このように構成要素を並べ替えた英文クレームを日本語に自動翻訳することにより、日本語への翻訳の精度が向上すると報告されている。

先行技術文献

特許文献

[0006] 特許文献1：特開２０１６－１６４７０７号公報

発明の概要

発明が解決しようとする課題

[0007] 特許文献１に記載された技術により、英文クレームを日本語クレームに精度高く翻訳することが可能になる。しかし、特許文献１に記載された技術では、英文クレームと日本語クレームとのパターンの作成及びそれらの構成要素の対応関係の特定はいずれも人手で行う必要があり、そのために自動翻訳の精度を向上させるためのコストが高いという問題がある。英文クレームを日本語クレームに翻訳する際のような、定型的な文章の翻訳をより精度高くかつ低コストで行うためには、上記したパターンの作成及び構成要素の対応関係をより低コストで行う必要がある。

[0008] それゆえに本発明は、定型的な文章を高い精度で自動翻訳するために、自

動翻訳に先立って、文の構成部品列を自動翻訳に適した順序に並べ替えるパターンを学習する自動翻訳パターン学習装置、そのパターンを用いて自動翻訳の前処理を行う前処理装置、及びそのためのコンピュータプログラムを提供することである。

課題を解決するための手段

[0009] 第１の局面に係る自動翻訳パターン学習装置は、第１の言語の文を第２の言語の文に自動翻訳により翻訳するために、第１の言語の文を複数個の構造部品に分解して順序を並べ替えるパターンを学習する。この自動翻訳パターン学習装置は、第１の言語及び第２の言語の対訳を記憶するための対訳記憶手段と、対訳記憶手段に記憶された対訳の各々について、第１の言語の文の構造部品の出現順序と、対応する第２の言語の文の構造部品の出現順序との対応関係が変化する位置を第１の言語の文及び第２の言語の文の各々の分割位置として検出するための検出手段と、第１の言語の文の、連続する所定個数の単語列と、当該単語列内の所定位置に分割位置があるか否かを示す正解情報を付したデータを機械学習の学習データとして生成するための学習データ生成手段と、学習データ生成手段により生成された学習データを用いて、機械学習により、第１の言語の文の所定個数の入力単語列が与えられたときに、当該入力単語列内の所定位置に分割位置があるか否かを判定するための判定モデルの学習を行うための学習手段とを含む。

[0010] 好ましくは、学習データ生成手段は、第１の言語の文の、連続する所定の偶数個数の単語列と、当該単語列内の中央に分割位置があるか否かを示す正解情報を付したデータを機械学習の学習データとして生成するための手段を含む。

[0011] より好ましくは、学習データ生成手段は、第１の言語の文の、連続する所定個数の単語列と、当該単語列内の所定位置に分割位置があるか否かを示す情報、及び分割位置の前後の単語列の各々について、第２の言語の文における当該単語が属する構成部品の並べ替えパターンを示す情報とからなる正解情報を付したデータを機械学習の学習データとして生成するための手段を含

む。

[0012] 第２の局面に係る前処理装置は、第１の言語の文を第２の言語の文に自動翻訳により翻訳する前処理として、第１の言語の文を複数個の構造部品列に分割し、その順序を並べ替える前処理装置であって、第１の言語の文の所定個数の入力単語列が与えられたときに、当該入力単語列内の所定位置に、第１の言語の文の分割位置があるか否かを判定するように、予め学習済の判定手段と、第１の言語の入力文が与えられたことに応答して、当該入力文を構成する、連続する所定個数の単語列を生成するための単語列生成手段と、単語列生成手段により生成された所定個数の単語列の各々を判定手段に与えることにより、入力文を構成する構成部品の分割位置を特定する分割位置特定手段と、分割位置特定手段により特定された分割位置で入力文を複数個の構造部品に分割し、判定手段に対応して定められる並べ替えパターンにしたがって、当該複数個の構成部品を並べ替える並べ替え手段とを含む。

[0013] 好ましくは、判定手段は、第１の言語の文の所定個数の入力単語列が与えられたときに、当該入力単語列内の所定位置に、第１の言語の文の分割位置があるか否かを判定し、当該入力単語列内の、分割位置前後の単語が属する構造部品が第２の言語の文のどの位置に配置されるかを示す分割位置情報を出力するように予め学習済の手段を含む。並べ替え手段は、分割位置特定手段により判定された分割位置で入力文を複数個の構造部品に分割し、判定手段により出力された分割位置情報にしたがって、当該複数個の構造部品を並べ替えるための手段を含む。

[0014] 第３の局面に係るコンピュータプログラムは、コンピュータを、上記したいずれかの装置の各手段として機能させる。

図面の簡単な説明

[0015] [図1]日英翻訳における構成部品列の並べ替えパターンの例を説明するための図である。

[図2]日英翻訳における構成部品列の並べ替えパターンの他の例を説明するための図である。

[図3]本発明の第１の実施の形態における分割ポイントの検出手法を説明するための模式図である。

[図4]本発明の第１の実施の形態における学習データの構成を説明する模式図である。

[図5]本発明の第１の実施の形態に係る自動翻訳システムの構成を示すブロック図である。

[図6]本発明の第１の実施の形態において、文の分割ポイントを検出するためのプログラムの擬似コードを示す図である。

[図7]図６に示す擬似コードにしたがうプログラムの動作を説明するための模式図である。

[図8]図５に示す自動翻訳システムにおいて、入力された文の分割を行うためのプログラムの制御構造を示すフローチャートである。

[図9]本発明の第２の実施の形態における学習データの構成を説明する模式図である。

[図10]本発明の各実施の形態に係る自動翻訳システム並びにそれを構成する学習装置及び前処理装置を実現するコンピュータの外観を示す図である。

[図11]図１０に示すコンピュータのハードウェア構成を示すブロック図である。

発明を実施するための形態

[0016] 以下の説明及び図面では、同一の部品には同一の参照番号を付してある。したがって、それらについての詳細な説明は繰返さない。

[0017] ［基本的考え方］

以下に本願発明の実施の形態を説明する前に、本願発明の基礎となっている基本的考え方を説明する。図３を参照して、「信号を発信できる通信装置を提供する」という日本語の文１１２と、「To provide a communication apparatus capable of sending signals」というその対訳文１１０との間の単語の対応関係を示す。日本語と英語との間では、一般的に語順が逆転するが、図３に示すように、両者の単語の対応関係は３個のブロック１１４，１１

６及び１１８に分けられる。第１のブロック１１４及び第３のブロック１１８では、英語の語順と日本語の語順とが互いに逆の関係になっている。これに対して第２のブロック１１６では、英語の語順と日本語の語順とが部分的ではあるが同一の関係になっている。図１及び図２と、図３とを比較すると、第１のブロック１１４、第２のブロック１１６，及び第３のブロック１１８を構成する日本語の単語列及び英語の単語列が、それぞれ一団の構成部品列となっていて、翻訳時にはこれらがそれぞれまとめて互いに並べ替えられる事がわかる。そして、その境界の１つは、第１のブロック１１４と第２のブロック１１６の境界（これを分割ポイントＢという。）であり、他の境界の１つは、第２のブロック１１６と第３のブロック１１８の境界（これを分割ポイントＡという。）であることが分かる。

[0018] 以下の実施の形態では、日本語と英語との間の対訳の、このような性質を利用し、入力される文の分割ポイントを推定し、その分割ポイントで分けられた構成部品列からなるブロックの順番を、翻訳のターゲット言語の順番に並べ替え、その後に自動翻訳を行う。こうすることで、ターゲット言語での翻訳が精度よく行えるようになる。

[0019] なお、以下の説明は日本語から英語への翻訳を例にしているが、言語の組み合わせはこれには限定されない。特に、語順が変わる事が多い、日本語と欧州言語との間の翻訳に本発明を適用できる。

[0020] ［第１の実施の形態］

＜構成＞

第１の実施の形態では、分割ポイントを推定するために統計的モデルを用いる。本実施の形態では統計的モデルとしてサポートベクターマシン（ＳＶＭ）を用いる。

[0021] ＳＶＭを用いるにあたって、その学習データが問題となる。本実施の形態では、自動翻訳装置を用いて対訳の間の単語対応表を求め、その情報を利用して図３に示すような単語の順序の間の関係を調べ、分割ポイントの位置を特定する。対訳の間の単語対応表は、高い信頼性で求められることが分かっ

ている。

[0022] 図４に、学習データの構成を示す。図４を参照して、日本語から英語に翻訳する際に、図３のような関係を利用して、入力文１１２の内の２個の分割ポイント１２０及び１２２を検出する。入力文内で連続する４個の単語からなる単語列を全て生成し、その中央に分割ポイントが存在するか否かを調べ、その情報を各単語列に付す。想定される分割ポイント位置に左側（前側）から隣接する単語列を左隣接単語群、右側（後側）から隣接する単語列を右隣接単語群という。図４では、これら各単語群に、その中央位置に分割ポイントがあるか否かを示す情報（タグ）を正解データとして付す。こうしてできたデータ１３０の各行が学習データとなる。

[0023] 図５を参照して、第１の実施の形態に係る自動翻訳システム１５０について、日本語を英語に翻訳する事例を説明する。自動翻訳システム１５０は、学習データの元となる対訳を記憶する記憶装置１６０と、記憶装置１６０に記憶された対訳の各々について、自動翻訳装置を用いて求めた単語対応表を各対訳と関連付けて記憶する記憶装置１６２と、記憶装置１６０に記憶された対訳及び記憶装置１６２に記憶された単語対応表を用いて、ＳＶＭからなる分割モデル１６６の学習を行うための学習処理部１６４とを含む。

[0024] 自動翻訳システム１５０はさらに、日本語の入力文１６８を受けて、分割モデル１６６を参照して、入力文１６８に対して図３に示すような並べ替えが可能であれば、入力文１６８を分割ポイントで３個のブロックに分割し、図３に示す対応関係にしたがって各ブロックの順番を英語の順番に並べ替える前処理を行う前処理部１７０と、前処理部１７０により前処理が行われた入力文１７１に対して、日英の自動翻訳を行って英語の出力文１７４を生成するための自動翻訳装置１７２とを含む。自動翻訳装置１７２としてはどのようなものを用いても良いが、本実施の形態では、後述するように、単語の事前並べ替えを行うものを採用すると特に高い精度で自動翻訳が行えることが分かった。

[0025] 学習処理部１６４は、記憶装置１６０に記憶された対訳の各々の日本語に

ついて、記憶装置１６２に記憶された対応する単語対応表を用いて分割ポイントを検出するための分割ポイント検出部１９０と、分割ポイント検出部１９０により分割ポイントが検出された各対訳の日本語の文から、図４に示す構成にしたがった学習データを生成するための学習データ生成部１９２と、学習データ生成部１９２により生成された学習データを記憶するための記憶装置１９４と、記憶装置１９４に記憶された学習データを用いて、ＳＶＭからなる分割モデル１６６の学習を行うためのモデル学習部１９６とを含む。

[0026] 前処理部１７０は、入力文１６８を受けて、入力文１６８から学習データと同じ単語数の、連続する単語からなる単語列を全て生成し、これら単語列の全てについて分割モデル１６６を参照することによりそれら単語列の中央に分割ポイントがあるか否かを判定し、入力文１６８に分割ポイントがあればその分割ポイントで分割する分割部２００と、分割部２００により分割されたブロックを、図１及び図２に示すような入替パターンにしたがって並べ替える処理を行って前処理後の入力文１７１を出力する並べ替え部２０２とを含む。

[0027] 図６は、図５に示す分割ポイント検出部１９０を実現するためのプログラムの制御構造を表す擬似コードであり、図７はその動作を説明するための図である。図６の擬似コードは、日本語の入力文の構造が図７に示すブロック１１８、ブロック１１６、及びブロック１１４の順番で、対応する英語の出力文の構造がブロック１１４、ブロック１１６、及びブロック１１８である場合に、日本語の入力文の２個の分割ポイントを検出する。入力文の構造が上記したものでない場合には、このプログラムは何も返さない。したがって図５に示す分割ポイント検出部１９０は、このプログラムから２個の分割ポイントが得られたときにはその分割位置を出力するが、何も得られない場合には分割ポイントは出力しない。この結果、分割ポイントが得られたときには処理対象の対訳から学習データが作成されるが、得られないときには処理対象の対訳は無視され、対応する学習データは作成されない。

[0028] 図６及び図７において、「align(i)」は、日本語の文のｉ番目の単語に対

応付けられた英訳文の単語位置を示す。このプログラムは、以下のような条件を満たすｉ、ｊの組を求めるものである。

[0029] （１）日本語の文の１番目の単語が英訳文の末尾の単語に対応し、それ以後、日本語の単語をｉ番目まで順番に進めると、対応する英語の単語は文の先頭に向かって順番に進み、

（２）ｉ＋１番目の日本語の単語に到達すると、対応する英語の単語が、先頭に向かっていくつかの単語を飛び越して進み、以後日本語の単語をｊ番目まで順番に進めると、対応する英語の単語は末尾に向かって逆順に進み、

（３）最後に、ｊ＋１番目の日本語の単語に到達すると、対応する英語の単語が、（２）でジャンプした単語よりさらに文の先頭に近い部分にジャンプし、以後、日本語の単語を文の末尾のＸ番目の単語まで順番に進めると、英語の単語は文の先頭に向かって順番に進む。

[0030] 以上から分かるように、このプログラムは、図１及び図２に示すような並べ替えパターンに適合する対訳について、その分割ポイントを検出するためのものである。図１及び図２に示した並べ替えパターンと異なる並べ替えパターンに適合する学習データを作成する場合には、図６に示したプログラムと同様の考え方を用いて、そのためのプログラムを作成する必要がある。

[0031] より具体的に、図７を参照しながら図６の擬似コードを説明する。日本語文の単語の総数をＸとする。図７において日本語の単語の位置をｘ座標、英語の単語の位置をｙ座標で表す。いずれも文頭の単語位置が１である。図７では、原点は左上にあり、ｘ座標は左から右に、ｙ座標は上から下に伸びる。ブロック１１８とブロック１１６との境界の左側の日本語の単語位置をｍとし、ブロック１１６とブロック１１４の境界の左側の単語位置をｎとする。ここでの処理は、日本語及び英語の対応関係から、図７に示すように英文クレームと日本語クレームとのパターンに応じて、入れ替えの単位と成る構成要素（ブロック）の境界となる条件を満たすｍ及びｎを求めることである。

[0032] 第３行目で計算されるmin1の値は次のように変化する。なお、以下では、

説明を簡明にするため、ｉ番目の日本語単語が各ブロック（例えばブロック１１６）内に属するものであることを、「ｉがブロック１１６内にある」と記載する。

[0033] （Ａ１）ｉがブロック１１８内にあるときにはalign(i)で表される英単語のｙ座標

（Ａ２）ｉがブロック１１６内にあるときにはalign(m)（図７では”capable”）で表される英単語のｙ座標

（Ａ３）ｉがブロック１１４内にあるときにはalign(i)で表される英単語のｙ座標

第４行で計算されるmin2の値は次のように変化する。

[0034] （Ｂ１）ｉ及びｊがいずれもブロック１１８内にあるときにはｉの値に関係なくalign(j)で表される英単語のｙ座標

（Ｂ２）ｊがブロック１１６内にあるときにはｉの値に関係なくalign(m+1)で表される英単語（図７では”a”）のｙ座標

（Ｂ３）ｊがブロック１１４内にあるときにはｉの値に関係なくalign(j)で表される英単語のｙ座標

図６の５行目は、ｉ＋１番目の単語からｊ番目の単語までに対応する英単語のｙ座標がいずれも１番目の日本語単語からｉ番目の日本語単語に対応する英単語のｙ座標より小さいことを要求している。この条件が満足されるのは、ｉがブロック１１８内にある場合とｉがブロック１１６内にあり、かつｊがブロック１１４内にある場合だけである。

[0035] 図６の６行目は、ｊ＋１番目の単語から最後の単語までに対応する英単語のｙ座標が、いずれもｉ＋１番目の単語からｊ番目の単語のｙ座標より小さいことを要求している。この条件が満足されるのは、ｉ及びｊがいずれもブロック１１８内にあるか、ｉがブロック１１８にあってかつｊがｎ以上であるか、ｉ及びｊがいずれもブロック１１４内にある場合に限られる。

[0036] 図６の５行目の条件と６行目の条件の論理積をとると、ｉ及びｊがいずれもブロック１１８内にあるか、ｉがブロック１１８にあってかつｊがｎであ

るか、のいずれかになる。

[0037] さらに、図６の７行目は、ｉ＋１番目の単語に対応する英単語のｙ座標が、ｊ番目の単語に対応する英単語のｙ座標より小さいことを要求している。この条件が満足されるのは、ｉ＜ｊであることを考慮すると、ｉ＋１及びｊがいずれもブロック１１６内にある場合に限られる。この条件と先の条件との論理積をとると、結局、ｉがブロック１１８内にあってかつｉ＋１はｍより大きく、かつｊがｎである、ということになる。ｉがブロック１１８内にあってかつｉ＋１がｍより大きくなるのはｉ＝ｍのときに限られる。したがって、こうした条件を満足するｉ、ｊを求めれば、ｉは入れ替えの対象となるブロック１１８の右端の日本語単語の位置ｍ、ｊは同じく入れ替えの対象となるブロック１１６の右端の日本語単語の位置ｎを表すことになる。

[0038] 図６の第５行から第７行は、上記した条件を吟味するための処理である。このようにしてｍとｎが見つかれば、それ以上処理を続ける必要はないので、この２個の値をそれぞれboundary1及びboundary2という戻り値として（第８行～第９行）このルーチンを終了し、このプログラムを呼び出した親ルーチンに制御を戻せば良い（第１０行）。

[0039] なお、図６に示した擬似コードはあくまで一例であって、これ以外にも上記した条件を満たすｍ及びｎを見つけ出すアルゴリズムであればどのようなものを採用しても良い。

[0040] 図８は、図５に示す分割部２００の機能を実現するプログラムの制御構造を示すフローチャートである。図８を参照して、このプログラムは、図５に示す日本語の入力文１６８の形態素解析を行うステップ２５０と、ステップ２５０の処理により得られた形態素列を用いて、図５に示す分割モデル１６６に入力する単語列を作成するステップ２５２とを含む。ステップ２５２で作成される単語列は、入力文１６８の内の、連続する４個の単語からなる全ての単語列である。

[0041] このプログラムはさらに、ステップ２５２で得られた全ての単語列を分割モデル１６６に与えることにより、各単語列の中央位置に分割ポイントがあ

るか否かの判定結果を得るステップ２５４と、ステップ２５４で２個の分割ポイントが得られたか否かを判定するステップ２５６と、ステップ２５６の判定で２個の分割ポイントが得られたことに応答して、その位置で入力文１６８を構造部品に分割し、所定の分割がされたことを示す信号とともに出力して処理を終わるステップ２５８と、ステップ２５６の判定で２個の分割ポイントが得られなかったことに応答して、さらにステップ２５６の判定で分割ポイントが３個あったか否かを判定するステップ２６０とを含む。

[0042] このプログラムはさらに、ステップ２６０の判定が肯定であるときに、翻訳前の言語がいわゆる主辞後置型言語（日本語、韓国語等）か否かを判定するステップ２６２と、ステップ２６２の判定が肯定であることに応答して、ステップ２６０の判定で得られた３個の分割ポイントのうち、翻訳前の文の文末側の２個の分割ポイントで入力文を分割し、３個ではなく２個の分割ポイントが得られたことを示す信号とともに出力して処理を終わるステップ２６４と、ステップ２６２の判定が否定であること、すなわち翻訳前の言語がいわゆる主辞前置型言語（英語、中国語等）であることに応答して、翻訳前の文の文頭側の２個の分割ポイントで入力文を分割し、３個ではなく２個の分割ポイントが得られたことを示す信号とともに出力して処理を終わるステップ２６６と、ステップ２６０の判定が否定であること、すなわち分割ポイントが１個以下、又は４個以上であることに応答して、２個の分割ポイントが得られなかったことを示す信号とともに入力文１６８をそのまま出力して処理を終わるステップ２６８とを含む。

[0043] ＜動作＞

以上に説明した自動翻訳システム１５０は以下のように動作する。図５を参照して、自動翻訳システム１５０の動作には、分割モデル１６６の学習と、分割モデル１６６を用いた入力文１６８の自動翻訳という２つのフェーズがある。これらを順番に説明する。

[0044] 〈分割モデル１６６の学習フェーズ〉

これに先立ち、記憶装置１６０には多数の対訳が記憶されているものとす

る。それら対訳の各々について、自動翻訳装置による対訳間の単語対応付がされ、単語対応表が作成された各対訳と関係付けて記憶装置１６２に記憶される。

[0045] 分割ポイント検出部１９０は、記憶装置１６０から順番に対訳を読出し、さらに記憶装置１６２から対応する単語対応表を読出し、図６に示す擬似コードで示されるような制御構造を持つプログラムによって、各対訳のうちの日本語の文の分割ポイントを特定する。なお、これら対訳は日本語と英語のように、いずれもブロック１１４、１１６、１１８が翻訳によりブロック１１８、１１６、１１４という順番に並べ替えられるような対訳であるものとする。分割ポイント検出部１９０は、各対訳データの日本語の文に、検出された２個の分割ポイントを特定する情報を付与して学習データ生成部１９２に出力する。

[0046] 学習データ生成部１９２は、分割ポイント検出部１９０から与えられた対訳の各々の日本語の文について、連続する４個の単語からなる単語列を全て生成し、その中央に分割ポイントがあるか否かを示す情報を各単語列に付与し記憶装置１９４に学習データとして蓄積する。

[0047] モデル学習部１９６は、記憶装置１９４に蓄積された学習データを用いて分割モデル１６６のトレーニングを行う。このトレーニングの結果、分割モデル１６６は、４個の単語からなる単語列が与えられると、その中央位置に分割ポイントがあるか否かを判定し、判定結果をその信頼度を示すスコアとともに出力するようにそのパラメータの値が調整される。

[0048] 全ての学習データによる分割モデル１６６のトレーニングが完了すると、そのトレーニング結果を用いた前処理部１７０及び自動翻訳装置１７２による自動翻訳を行うことができる。

[0049] 〈自動翻訳フェーズ〉

図５を参照して、自動翻訳システム１５０による自動翻訳は以下のようにして行われる。日本語の入力文１６８が与えられると、分割部２００は入力文１６８を形態素解析し（図８のステップ２５０）、その結果を用いて、入

力文１６８の中で連続する４個の単語からなる単語列を全て生成する（ステップ２５２）。次いで分割部２００は、生成された各単語列を分割モデル１６６への入力とし、当該単語列内の分割ポイントの有無に関する分割モデル１６６の判定結果を、上述した信頼度を示すスコアとともに受け取る（ステップ２５４）。ステップ２５４の処理により、全ての単語列の内に２個のみ分割ポイントがあるという判定結果が得られると、分割部２００は、その分割ポイントで入力文１６８を分割し、２個の分割ポイントが得られたという信号とともに並べ替え部２０２に出力する（ステップ２５８）。もしも分割ポイントが１個以下、又は４個以上の場合には、分割部２００は入力文１６８をそのまま出力する（ステップ２６８）。分割部２００はこのとき、２個の分割ポイントが得られなかったことを示す信号を並べ替え部２０２に出力する。さらに、分割ポイントが３個あったときには、翻訳前の言語が主辞後置型言語であれば翻訳前の文の文末側の２個の分割ポイントで文を分割して出力し（ステップ２６４）、翻訳前の言語が主辞前置型言語であれば文頭側の２個の分割ポイントで文を分割して出力する（ステップ２６６）。このとき、ステップ２６４とステップ２６６のいずれにおいても、２個の分割ポイントが得られたという信号が並べ替え部２０２に出力される。

[0050] 並べ替え部２０２は、分割部２００から２個の分割ポイントが得られたという信号を受け取ったときには、分割モデル１６６により予め定められた並べ替えパターンにしたがって、分割部２００から受信した３個のブロックを並べ替え、入力文１７１として自動翻訳装置１７２に与える。もしも２個の分割ポイントが得られなかったという信号を受け取ったときには、並べ替え部２０２は分割部２００の出力をそのまま入力文１７１として自動翻訳装置１７２に与える。

[0051] 自動翻訳装置１７２は、入力文１７１を入力として、日英の翻訳を行い、英語の出力文１７４を出力する。分割部２００による分割と並べ替え部２０２による並べ替えとが行われた場合、自動翻訳装置１７２による翻訳では、入力文が出力である英語の語順に近い形となる。その結果、翻訳の精度が高

まることが知られている。以下に説明するように、実験でもそれが確認できた。

[0052] ＜実験結果＞

実験では、日本の公開特許公報の英文抄録（ＰＡＪ）と、もとになった日本語抄録とについて、自動的に単語対応付けを行った対訳コーパスを用いた。統計的機械翻訳の学習用にこれらのうち１，０００，０００文を使用し、開発・テスト用にそれぞれ１，０００文を使用した。この対訳コーパスのうち１００，０００文を対象に、上記した並べ替えの必要な３８，１９４文を抽出し、これらから学習データを作成して分割モデル１６６としてＳＶＭのトレーニングを行った。なお、以下の表では、上記実施の形態でのブロックごとの並べ替えを「グローバルな並べ替え」と呼び、以前から統計的機械翻訳で用いられている、単語レベルでの並べ替えを「単語並べ替え」と呼んでいる。

[0053] 自動翻訳は以下の４種類の設定で行い、翻訳結果はRIBES値により評価した。RIBES値は単語の並べ替えの情報を重視する評価尺度として知られている。チェックマークはその列の見出しに記載された手法を採用したことを示す。

[0054] [表1]

	グローバルな並べ替え	単語並べ替え	RIBES値
T1			44.9
T2	✓		61.0
T3		✓	64.9
T4	✓	✓	72.1

この表から分かるとおり、グローバルな並べ替えを用いると（Ｔ２）、並べ替えの手法を何も採用しない場合（Ｔ１）と比較してRIBES値が１６．１ポイント上昇した。これは、単語並べ替えを単独で用いた場合（Ｔ３）のRIBES値の２０ポイント上昇という数値という値には及ばないが、それでも精度の改善値としては大きな値である。さらに、グローバルな並べ替えを単語並べ替

えと併用する（Ｔ４）と、単語並べ替えを単独で採用した場合（Ｔ３）と比較してさらに７．２ポイントと、精度を大幅に改善できた。したがって、この手法を単語並べ替えと併用することにより、日英翻訳での精度を大きく改善できることが分かった。翻訳の方向としては逆でも同様の結果が得られる。また、日英の組み合わせ以外にも、任意の言語の組み合わせに適用可能であり、特に語順が互いに大きく異なる言語の組み合わせに適用すると特に効果的である。

[0055] なお、この実施の形態に係る自動翻訳システム１５０で使用する分割モデル１６６は、図１及び図２に示されるような並べ替えパターンに対してのみ有効である。しかし、これ以外にも並べ替えパターンはあり得る。そうした並べ替えパターンについても同様の手法を適用可能であり、学習データを変えればよいことは容易に理解可能である。さらに、そのように異なる分割モデルを併用し、同じ入力文に対してそれら分割モデルを別々に適用することで、種々の並べ替えを行って翻訳することが可能になる。この場合には、複数の分割モデルの出力する分割ポイントのスコアを比較し、最も高いスコアを与えた分割モデルの判定を用いることが有効と考えられる。分割モデルによる分割数が２個以下及び４個以上の場合についても同様である。

[0056] ［第２の実施の形態］

上記第１の実施の形態では、１個の分割モデルによりブロックの並べ替え方法は一通りに定められる。しかし本発明はそのような実施の形態には限定されない。例えば、１個の分割モデルにより分割個数が同じ複数通りの並べ替えパターンを推定することも可能である。そのためには、学習データとして、分割個数が同じで並べ替えパターンが互いに異なる対訳を多数集めることが必要である。これら対訳から学習データを作成し、分割モデルの学習を行う点では第１の実施の形態と同様である。問題は、学習データの形式である。

[0057] 第１の実施の形態では、各単語列の中央位置に分割ポイントがあるか否かのみを判定していた。しかし、分割モデルはそのようなものには限定されな

い。例えば、図９に示すような構成の学習データを準備することにより、１個の分割モデルで同じ分割数で複数通りの並べ替えパターンを特定できる。

[0058] 図９を参照して、文１１２は分割ポイント１２０及び１２２により３個のブロック３００、３０２、及び３０４に分割される。英語に翻訳すると、これらはブロック３０４、３０２、及び３００の順番に並べ替えられる。このような条件で作成されるのが学習データ３３０である。学習データ３３０の構成が図４に示すものと異なるのは、単純に単語列の中央に分割ポイントがあるか否かを示す情報ではなく、左隣接単語群と右隣接単語群がそれぞれ並べ替え後にどのブロックに属するかを示す情報であるブロック番号ペア欄３４０を持つことである。このブロック番号ペアが同じ数値なら（例えば「３，３」）、左隣接単語群と右隣接単語群とは、並べ替え後も同じブロックに属する、すなわちこの単語列は分割ポイントを持たない事がわかる。また、ブロック番号ペアが異なる数値、例えば「３，２」であれば、左隣接単語群と右隣接単語群とが、並べ替え後にそれぞれ３番目と２番目のブロックに分けられることが分かる。すなわち、この単語列は分割ポイントを持つことが分かり、しかもブロック番号ペアを見ることによりそれぞれ並べ替え後にどの位置に移動するかも分かる。なお、この場合も分割ポイントは４個の単語の中央位置にあることを想定している。

[0059] こうした学習データを多数準備して分割モデルをトレーニングし、各単語列について並べ替え後のブロック番号ペアを推定することで、１個の分割モデルを用いて分割ポイントの位置と並べ替えパターンとを推定できる。

[0060] ［コンピュータによる実現］

本発明の実施の形態に係る自動翻訳システム１５０及びその構成要素は、いずれもコンピュータハードウェアと、そのコンピュータハードウェア上で実行されるコンピュータプログラムとにより実現できる。図１０はこのコンピュータシステム６３０の外観を示し、図１１はコンピュータシステム６３０の内部構成を示す。

[0061] 図１０を参照して、コンピュータシステム６３０は、メモリポート６５２

及びDVD（Digital Versatile Disk）ドライブ６５０を有するコンピュータ６４０と、いずれもコンピュータ６４０に接続されたキーボード６４６と、マウス６４８と、モニタ６４２とを含む。

[0062] 図１１を参照して、コンピュータ６４０は、メモリポート６５２及びDVDドライブ６５０に加えて、CPU（中央処理装置）６５６と、CPU６５６、メモリポート６５２及びDVDドライブ６５０に接続されたバス６６６と、起動プログラム等を記憶する読出専用メモリ（ROM）６５８と、バス６６６に接続され、上記自動翻訳システム１５０の各部の機能を実現するプログラム命令、システムプログラム及び作業データ等を記憶するランダムアクセスメモリ（RAM）６６０と、ハードディスク６５４を含む。コンピュータシステム６３０はさらに、他端末との通信を可能とするネットワーク６６８への接続を提供するネットワークインターフェイス（I/F）６４４を含む。

[0063] コンピュータシステム６３０を上記した実施の形態に係る自動翻訳システム１５０及びその各機能部として機能させるためのコンピュータプログラムは、DVDドライブ６５０又はメモリポート６５２に装着されるDVD６６２又はリムーバブルメモリ６６４に記憶され、さらにハードディスク６５４に転送される。又は、プログラムはネットワーク６６８を通じてコンピュータ６４０に送信されハードディスク６５４に記憶されてもよい。プログラムは実行の際にRAM６６０にロードされる。DVD６６２から、リムーバブルメモリ６６４から又はネットワーク６６８を介して、直接にRAM６６０にプログラムをロードしてもよい。

[0064] このプログラムは、コンピュータ６４０を、上記実施の形態に係る自動翻訳システム１５０の各機能部として機能させるための複数の命令からなる命令列を含む。コンピュータ６４０にこの動作を行わせるのに必要な基本的機能のいくつかはコンピュータ６４０上で動作するオペレーティングシステム若しくはサードパーティのプログラム又はコンピュータ６４０にインストールされる、ダイナミックリンク可能な各種プログラミングツールキット又はプログラムライブラリにより提供される。したがって、このプログラム自体

はこの実施の形態のシステム、装置及び方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能又はプログラミングツールキット又はプログラムライブラリ内の適切なプログラムを実行時に動的に呼出すことにより、上記したシステム、装置又は方法としての機能を実現する命令のみを含んでいればよい。もちろん、独立したプログラムのみで必要な機能を全て提供してもよい。

[0065] 以上、実施の形態を例として本願発明を説明したが、本願発明が上記実施の形態に限定されるわけではない。例えば、既に述べたように、複数種類の分割モデルを並列に用い、入力文に対して最も高いスコアで分割ポイントを検出した分割モデルの結果にしたがって入力文を分割し、並べ替えるようにしてもよい。また、自動翻訳装置１７２として、上記実験では単語並べ替えを用いた場合に最もよい精度が得られることが分かったが、自動翻訳の仕組みが単語並べ替えを用いたものに限定されるわけではない。統計的な言語モデル又は翻訳モデルを用いて翻訳を行う自動翻訳装置であればどのようなものでも用いることができる。なお、この場合の分割モデルの学習データは、前述したとおり、並べ替えパターンにしたがった対訳のみを学習データとして選択してその分割ポイントを検出するようなプログラムにより行う必要がある。そうしたプログラムは、並べ替えパターンに伴う単語の対応付けを検討し、その特徴を捉えるようにすることで作成できる。

[0066] また、上記実施の形態では、分割モデルに入力する単語列の長さを４としたが、この値に限定されるわけではない。単語列は２以上であればどのような長さでもよい。また、分割ポイントの有無を判定する位置は単語列の中央としたが、それには限定されない。単語列の長さが奇数の場合も含め、分割ポイントの位置が中央以外で単語列内の単語の間であればどのような位置でもよい。もっとも、おそらくは分割ポイントの位置は中央又はその前後が最も望ましいと考えられる。

[0067] 上記実施の形態では、分割モデルとしてＳＶＭを用いたが、使用するモデ

ルがＳＶＭに限定されるわけではない。例えばニューラルネットワークを用いてもよい。

[0068] また、上記実施の形態では、翻訳言語のペアとして日本語と英語とを扱った。しかし、本発明はそのような実施の形態には限定されず、任意の言語ペアに対して適用できる。ただし、日本語と英語、日本語と中国語等のように、互いの語順が大きく異なる言語ペアの間に本発明を適用することにより、語順が似通った言語ペアより大きな効果を得ることができる。

産業上の利用可能性

[0069] 本発明は、いわゆる翻訳産業に限らず、利用者が母語以外の言語に遭遇する可能性がある産業全般に利用可能である。例えば製造業において外国から装置、製品又は原料を輸入するための手続きを行う産業、そうした装置、製品又は原料を目的地に正しく配送する産業、配送された装置及び製品を正しくセットアップし、原料を正しく利用することが必要な産業等、幅広い産業において本発明は利用可能である。

[0070] 今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内での全ての変更を含む。

符号の説明

[0071] １１４　第１のブロック

１１６　第２のブロック

１１８　第３のブロック

１２０、１２２　分割ポイント

１５０　自動翻訳システム

１６０、１６２、１９４　記憶装置

１６４　学習処理部

１６６　分割モデル

１６８、１７１　入力文

１７０　前処理部

１７２　自動翻訳装置

１９０　分割ポイント検出部

１９２　学習データ生成部

１９６　モデル学習部

２００　分割部

２０２　並べ替え部

請求の範囲

[請求項1] 第１の言語の文を第２の言語の文に自動翻訳により翻訳するために、前記第１の言語の文を複数個の構造部品に分解して順序を並べ替えるパターンを学習する自動翻訳パターン学習装置であって、

前記第１の言語及び前記第２の言語の対訳を記憶するための対訳記憶手段と、

前記対訳記憶手段に記憶された対訳の各々について、前記第１の言語の文の構造部品の出現順序と、対応する前記第２の言語の文の構造部品の出現順序との対応関係が変化する位置を前記第１の言語の文及び前記第２の言語の文の各々の分割位置として検出するための検出手段と、

前記第１の言語の文の、連続する所定個数の単語列と、当該単語列内の所定位置に前記分割位置があるか否かを示す正解情報を付したデータを機械学習の学習データとして生成するための学習データ生成手段と、

前記学習データ生成手段により生成された学習データを用いて、機械学習により、前記第１の言語の文の前記所定個数の入力単語列が与えられたときに、当該入力単語列内の前記所定位置に前記分割位置があるか否かを判定するための判定モデルの学習を行うための学習手段とを含む、自動翻訳パターン学習装置。

[請求項2] 前記学習データ生成手段は、前記第１の言語の文の、連続する所定の偶数個数の単語列と、当該単語列内の中央に前記分割位置があるか否かを示す正解情報を付したデータを前記機械学習の学習データとして生成するための手段を含む、請求項１に記載の自動翻訳パターン学習装置。

[請求項3] 前記学習データ生成手段は、前記第１の言語の文の、連続する所定個数の単語列と、当該単語列内の所定位置に前記分割位置があるか否かを示す情報、及び前記分割位置の前後の単語列の各々について、前記

第２の言語の文における当該単語が属する構成部品の並べ替えパターンを示す情報とからなる正解情報を付したデータを機械学習の学習データとして生成するための手段を含む、請求項１に記載の自動翻訳パターン学習装置。

[請求項4] 第１の言語の文を第２の言語の文に自動翻訳により翻訳する前処理として、前記第１の言語の文を複数個の構造部品列に分割し、その順序を並べ替える前処理装置であって、

前記第１の言語の文の所定個数の入力単語列が与えられたときに、当該入力単語列内の所定位置に、前記第１の言語の文の分割位置があるか否かを判定するように、予め学習済の判定手段と、

前記第１の言語の入力文が与えられたことに応答して、当該入力文を構成する、連続する前記所定個数の単語列を生成するための単語列生成手段と、

前記単語列生成手段により生成された前記所定個数の単語列の各々を前記判定手段に与えることにより、前記入力文を構成する構成部品の分割位置を特定する分割位置特定手段と、

前記分割位置特定手段により特定された分割位置で前記入力文を複数個の構造部品に分割し、前記判定手段に対応して定められる並べ替えパターンにしたがって、当該複数個の構成部品を並べ替える並べ替え手段とを含む、前処理装置。

[請求項5] 前記判定手段は、

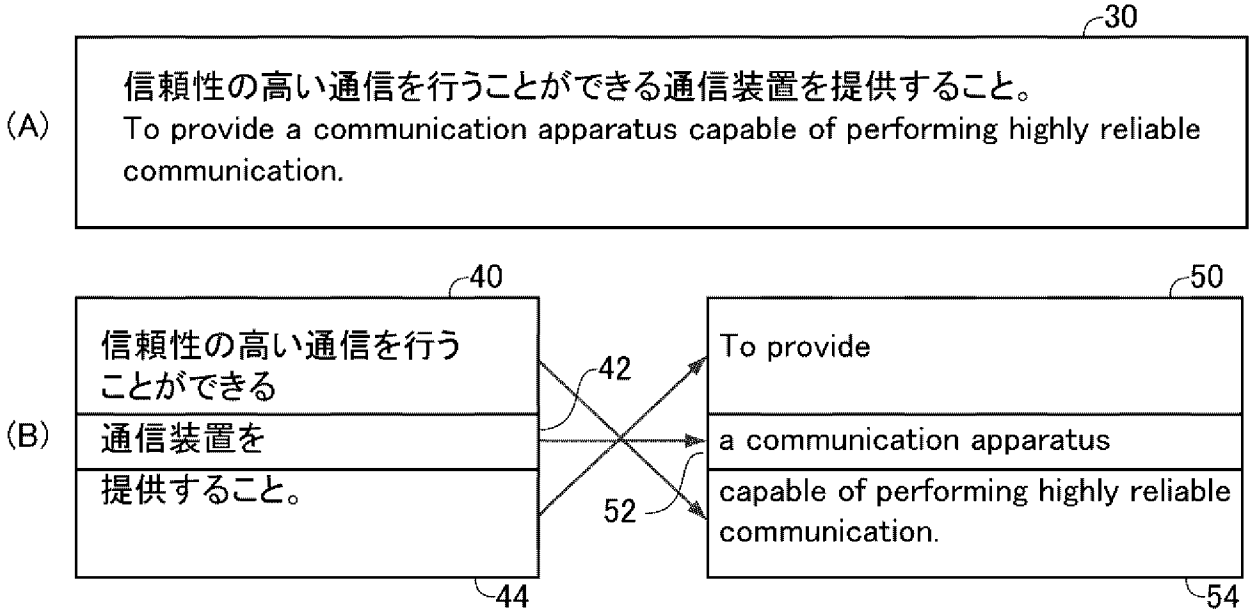
前記第１の言語の文の所定個数の入力単語列が与えられたときに、当該入力単語列内の所定位置に、前記第１の言語の文の分割位置があるか否かを判定し、当該入力単語列内の、前記分割位置前後の単語が属する構造部品が前記第２の言語の文のどの位置に配置されるかを示す分割位置情報を出力するように予め学習済の手段を含み、

前記並べ替え手段は、前記分割位置特定手段により判定された分割位置で前記入力文を複数個の構造部品に分割し、前記判定手段により

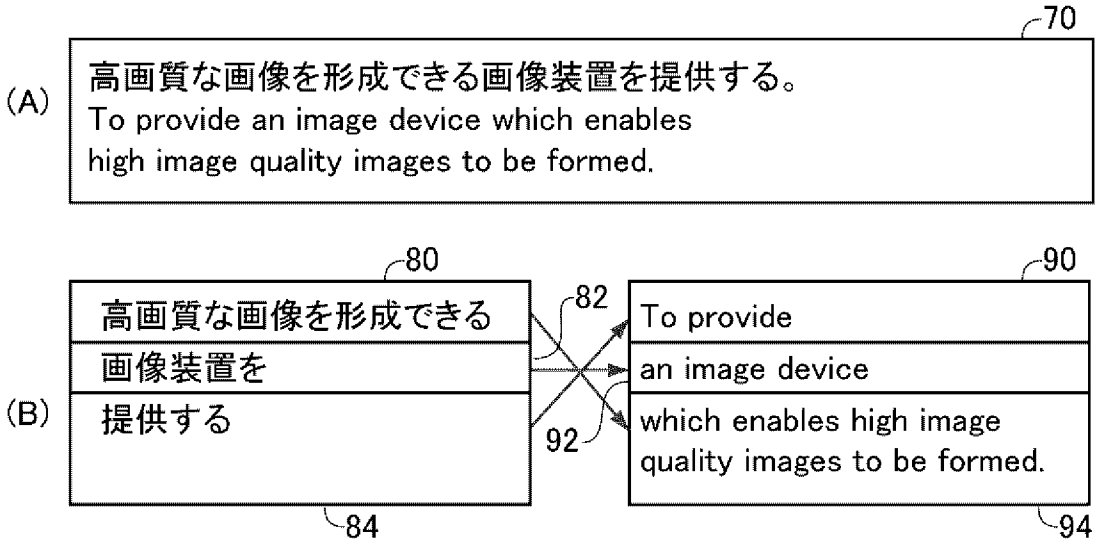
出力された前記分割位置情報にしたがって、当該複数個の構造部品を並べ替えるための手段を含む、請求項４に記載の前処理装置。

[請求項6] コンピュータを、請求項１～請求項５のいずれかに記載の各手段として動作させる、コンピュータプログラム。

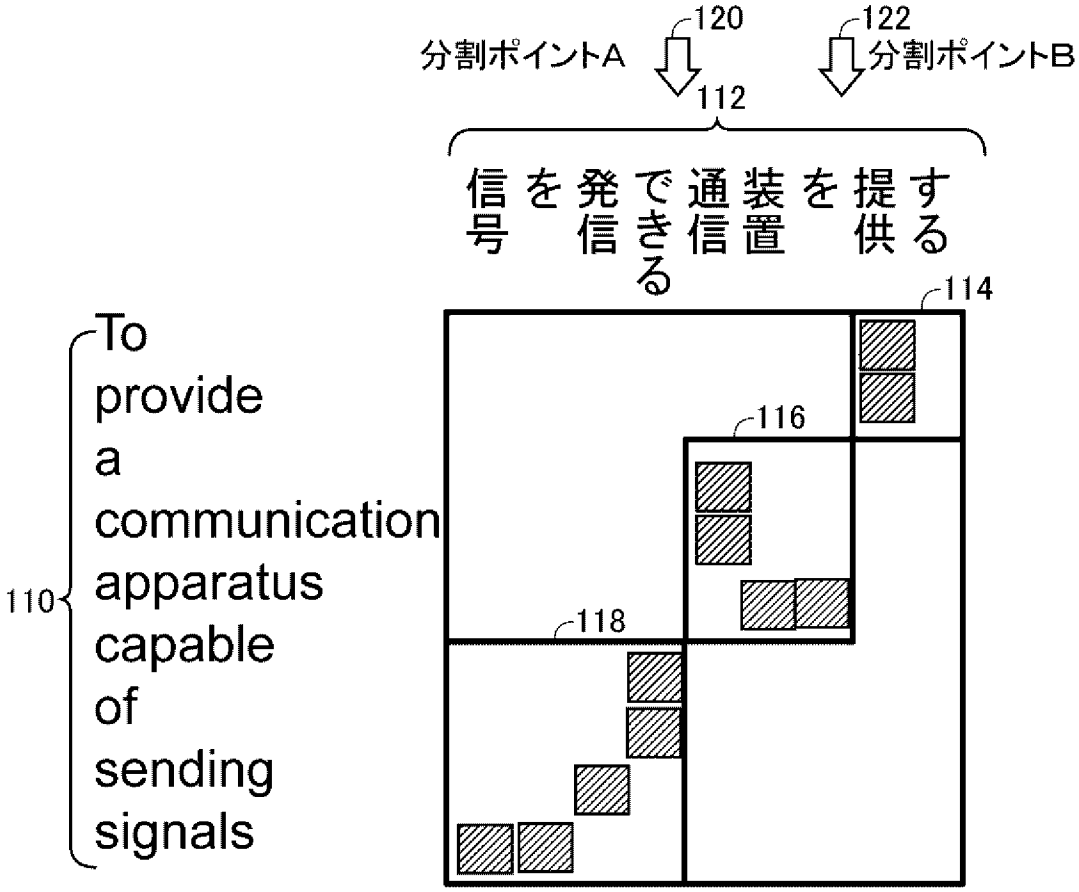
[図1]


(A)
30
信頼性の高い通信を行うことができる通信装置を提供すること。
To provide a communication apparatus capable of performing highly reliable communication.
(B)
40
信頼性の高い通信を行う
ことができる
通信装置を
提供すること。
44
42
52
50
To provide
a communication apparatus
capable of performing highly reliable communication.
54

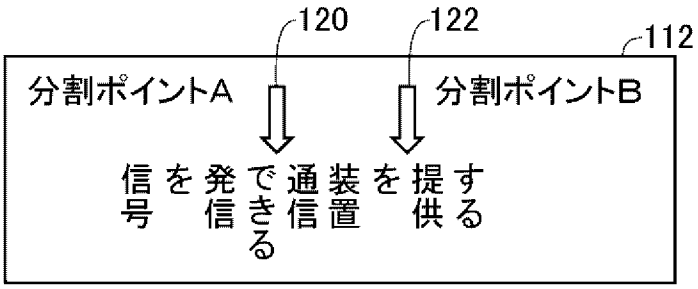
[図2]


(A)
70
高画質な画像を形成できる画像装置を提供する。
To provide an image device which enables high image quality images to be formed.
(B)
80
高画質な画像を形成できる
画像装置を
提供する
84
82
92
90
To provide
an image device
which enables high image quality images to be formed.
94

[図3]



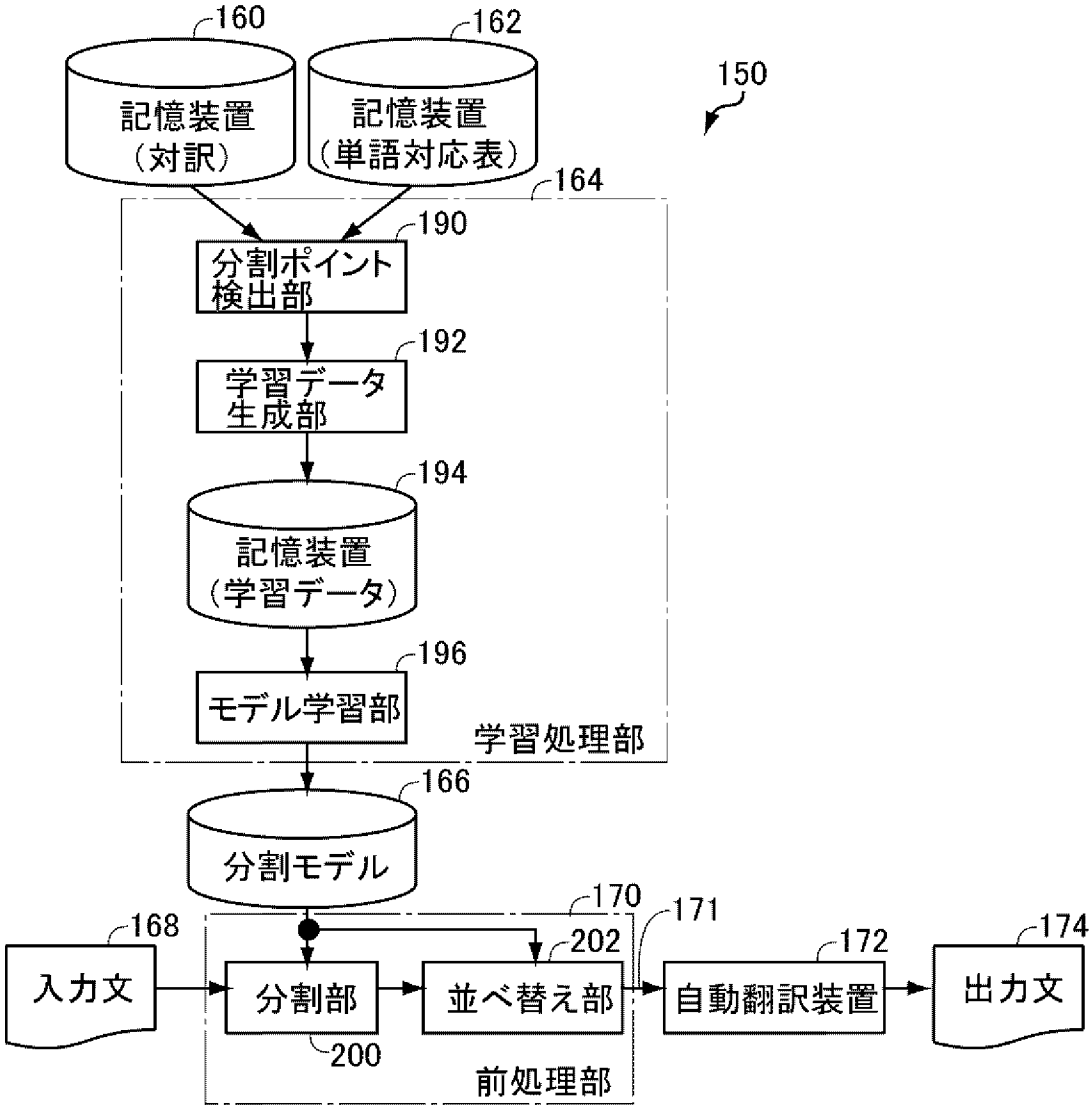
[図4]



130

学習器への入力		正解データ
左隣接単語群	右隣接単語群	分割する・しない
信号 を	発信 できる	×
を 発信	できる 通信	×
発信 できる	通信 装置	○
できる 通信	装置 を	×
通信 装置	を 提供	×
装置 を	提供 する	○

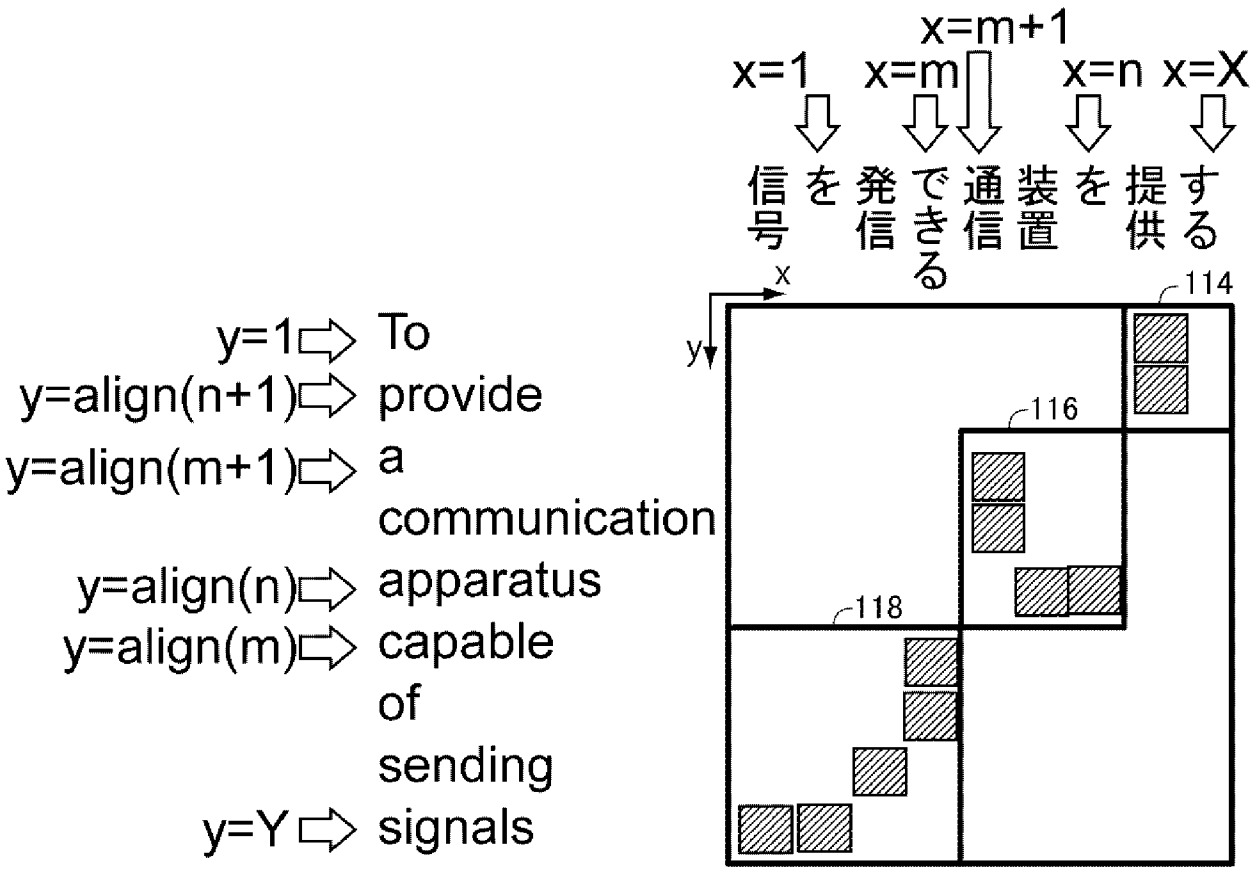
[図5]



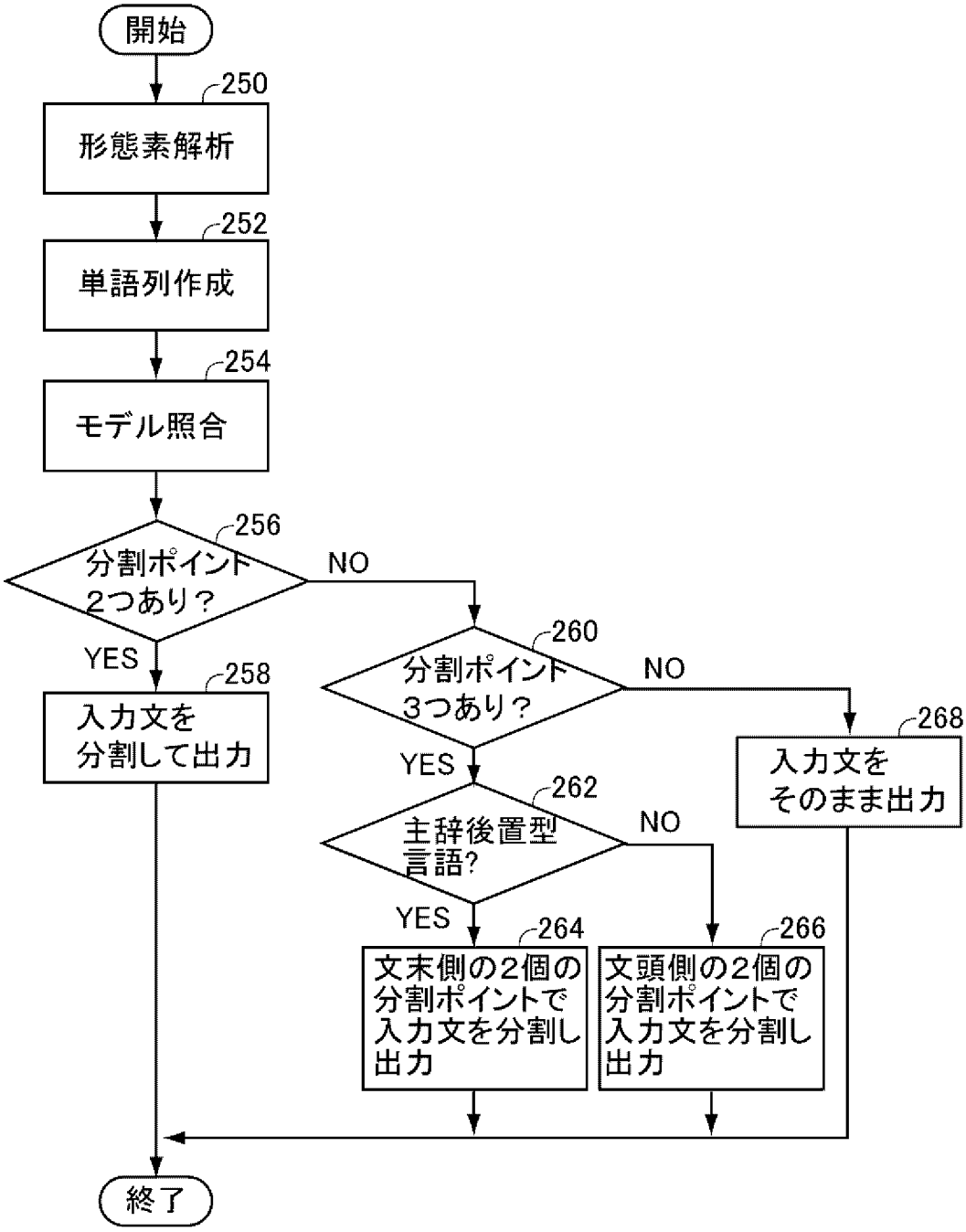
[図6]

```
001:  for j = 1 to X
002:        for i = 1 to j-1
003:              min1 = min y in [align(1)...align(i)]
004:              min2 = min y in [align(i+1)...align(j)]
005:              if(align(i+1)...align(j) <min1 and
006:                 align(j+1)...align(X)<min2 and
007:                 align(i+1)<align(j)) then
008:                      boundary1 = i
009:                      boundary2 = j
010:                      return boundary1 and boundary2
011:              end if
012:        end for
013:  end for
```

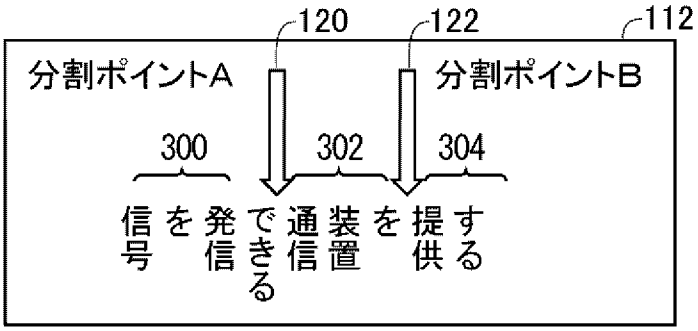
[図7]


x=m+1
x=1
x=m
x=n
x=X
信号を発信できる通信装置を提供する
x
y
114
116
118
y=1
To
y=align(n+1)
provide
y=align(m+1)
a
communication
y=align(n)
apparatus
y=align(m)
capable
of
sending
y=Y
signals

[図8]


開始
250
形態素解析
252
単語列作成
254
モデル照合
256
分割ポイント
2つあり？
NO
YES
258
入力文を
分割して出力
260
分割ポイント
3つあり？
NO
YES
262
主辞後置型
言語?
NO
YES
264
文末側の2個の
分割ポイントで
入力文を分割し
出力
266
文頭側の2個の
分割ポイントで
入力文を分割し
出力
268
入力文を
そのまま出力
終了

[図9]

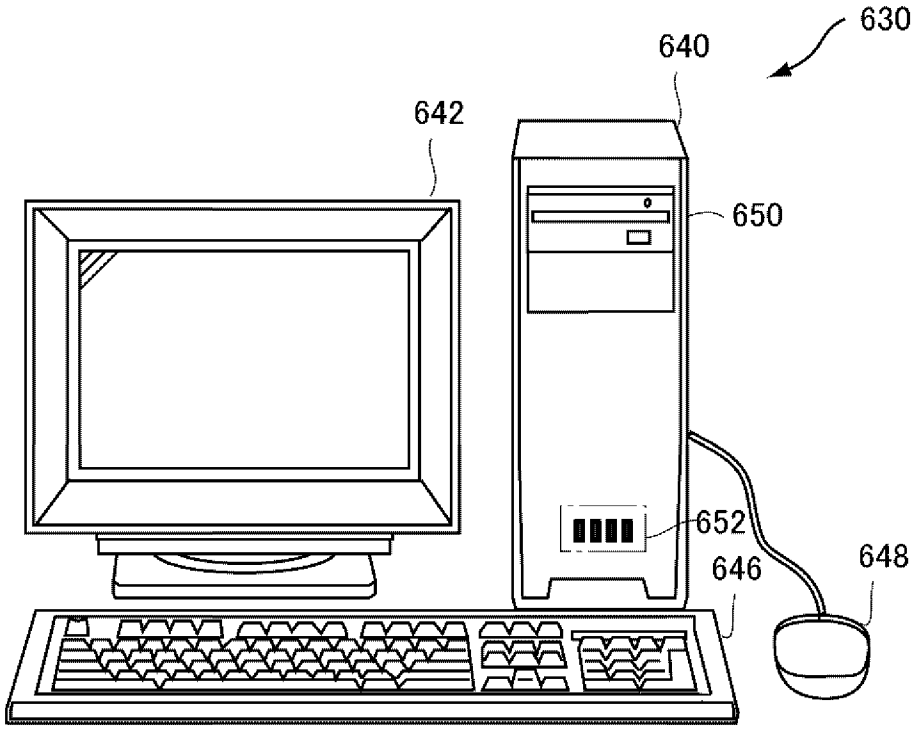


330

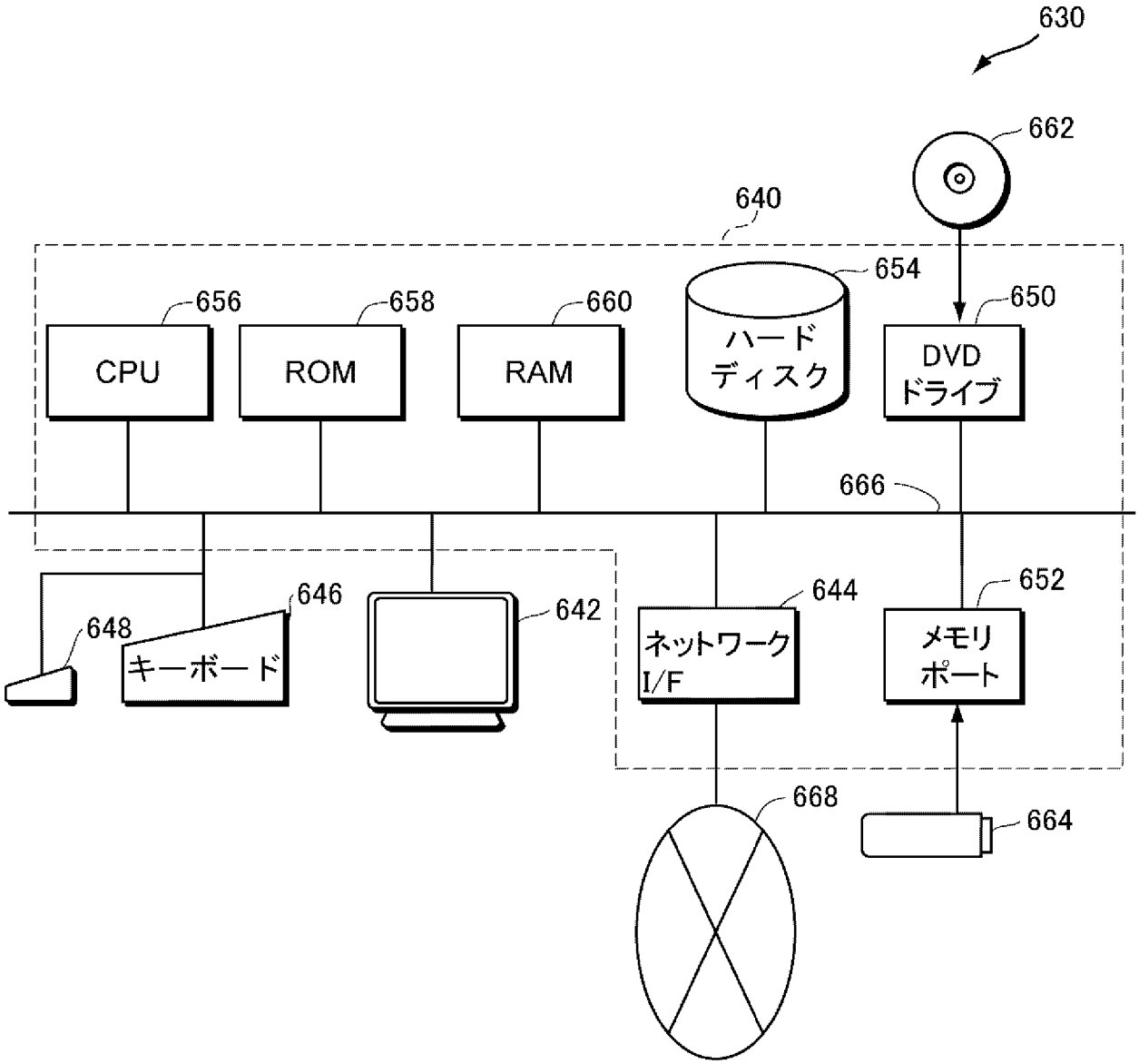
学習器への入力		学習器への正解
左隣接単語群	右隣接単語群	ブロック番号ペア
信号 を	発信 できる	3, 3
を 発信	できる 通信	3, 3
発信 できる	通信 装置	3, 2
できる 通信	装置 を	2, 2
通信 装置	を 提供	2, 2
装置 を	提供 する	2, 1

340

[図10]



[図11]


630
662
640
654
656
658
660
650
CPU
ROM
RAM
ハード
ディスク
DVD
ドライブ
666
646
642
644
652
648
キーボード
ネットワーク
I/F
メモリ
ポート
668
664

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2017/041249

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl. G06F17/28(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl. G06F17/27-17/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2018
Registered utility model specifications of Japan	1996-2018
Published registered utility model applications of Japan	1994-2018

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	渡辺太郎, 統計的機械翻訳の現場, 人工知能学会誌, 01 May 2012, vol. 27 no. 3, pp. 288-295, (WATANABE, Taro, Field of Statistical Machine Translation, Journal of Japanese Society for Artificial Intelligence)	1-6
A	JP 2015-153182 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 24 August 2015, entire text, all drawings (Family: none)	1-6

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 January 2018 (30.01.2018)	13 February 2018 (13.02.2018)

Name and mailing address of the ISA/	Authorized officer
Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Telephone No.

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１７／０４１２４９

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G06F17/28(2006.01)i

Ｂ．　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl.　G06F17/27-17/28

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１８年
日本国実用新案登録公報	１９９６－２０１８年
日本国登録実用新案公報	１９９４－２０１８年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

Ｃ．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
A	渡辺太郎，統計的機械翻訳の現場，人工知能学会誌，2012.05.01，Vol.27 No.3，288-295 頁	1-6
A	JP 2015-153182 A（日本電信電話株式会社）2015.08.24，全文・全図（ファミリーなし）	1-6

☐　Ｃ欄の続きにも文献が列挙されている。

☐　パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「＆」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
３０．０１．２０１８	１３．０２．２０１８

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
東京都千代田区霞が関三丁目４番３号

特許庁審査官（権限のある職員）　５Ｍ　９１９２
成瀬　博之
電話番号　０３－３５８１－１１０１　内線　３５９９