

(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
2017 年 1 月 5 日(05.01.2017)



(10) 国際公開番号
WO 2017/002199 A1

- (51) 国際特許分類:
G06F 17/27 (2006.01)
- (21) 国際出願番号: PCT/JP2015/068828
- (22) 国際出願日: 2015 年 6 月 30 日(30.06.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 楽天株式会社(RAKUTEN, INC.) [JP/JP]; 〒1580094 東京都世田谷区玉川一丁目 1 4 番 1 号 Tokyo (JP).
- (72) 発明者: 江木 聡志(EGI Satoshi); 〒1580094 東京都世田谷区玉川一丁目 1 4 番 1 号 楽天株式会社内 Tokyo (JP).
- (74) 代理人: 長谷川 芳樹, 外(HASEGAWA Yoshiki et al.); 〒1000005 東京都千代田区丸の内二丁目 1 番 1 号丸の内 MY PLAZA (明治安田生命ビル) 9 階 創英国際特許法律事務所 Tokyo (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, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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.

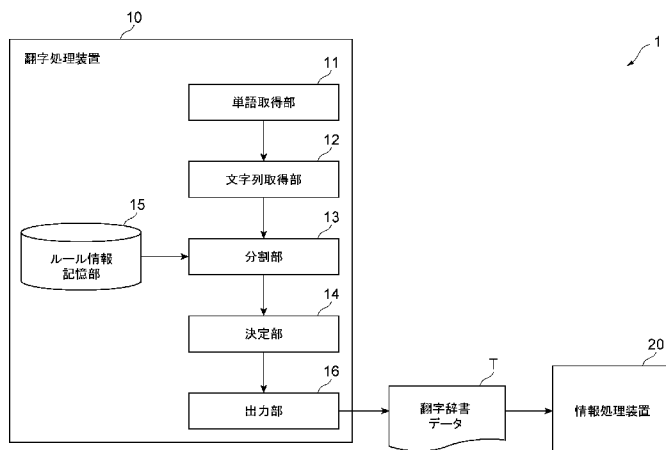
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,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

— 国際調査報告 (条約第 21 条(3))

(54) Title: TRANSLITERATION PROCESSING DEVICE, TRANSLITERATION PROCESSING METHOD, TRANSLITERATION PROCESSING PROGRAM AND INFORMATION PROCESSING DEVICE

(54) 発明の名称: 翻字処理装置、翻字処理方法、翻字処理プログラム、及び情報処理装置



- 10 Transliteration processing device
11 Word acquisition unit
12 Character string acquisition unit
13 Dividing unit
14 Determination unit
15 Rule information storage unit
16 Output unit
20 Information processing device
T Transliteration dictionary data

(57) Abstract: A transliteration processing device as in one embodiment of the present invention is provided with: a character string acquisition unit which acquires a first alphabetical character string in which a first word written in a first language having a prescribed character system is expressed by an alphabetical notation, and a second alphabetical character string in which a second word written in a second language having a character system that differs from that of the first language is expressed by an alphabetical notation; a determination unit which executes a determination of whether a first consonant element included in the first alphabetical character string and a second consonant element included in the second alphabetical character string are in a predetermined correspondence relationship, and on the basis of the determination result, determines whether the first word and the second word are in a transliteration relationship; and an output unit which outputs, as a pair, the first word and the second word which were determined by the determination unit to be in a transliteration relationship.

(57) 要約:

[続葉有]

一実施形態に係る翻字処理装置は、所定の文字体系を有する第１の言語で記述された第１の単語をアルファベット表記で表した第１のアルファベット文字列と第１の言語とは異なる文字体系を有する第２の言語で記述された第２の単語をアルファベット表記で表した第２のアルファベット文字列とを取得する文字列取得部と、第１のアルファベット文字列に含まれる第１の子音要素と第２のアルファベット文字列に含まれる第２の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて第１の単語と第２の単語とが翻字関係にあるか否かを決定する決定部と、決定部によって翻字関係にあると決定された第１の単語及び第２の単語を翻字対として出力する出力部と、を備える。

明 細 書

発明の名称：

翻字処理装置、翻字処理方法、翻字処理プログラム、及び情報処理装置
技術分野

[0001] 本発明の一側面は、翻字処理装置、翻字処理方法、翻字処理プログラム、及び情報処理装置に関する。

背景技術

[0002] 従来、一の言語（例えば日本語）で記述された単語と、当該単語と翻字関係にある他の言語（例えば英語）で記述された単語とのペア（以下「翻字対」）が登録された翻字辞書データが、種々の用途で利用される。例えば、翻字辞書データは、インターネット上で商品（サービスを含む）を販売するウェブサイト（ＥＣサイト）において、ユーザによって入力されるキーワードに基づく商品検索を行う際等に利用される。

[0003] 例えば上記の商品検索の際に、当該商品検索を実行するコンピュータは、翻字辞書データを参照することにより、日本語のカタカナ表記の「ボジョレー」が検索キーワードとして入力された場合に、日本語のカタカナ表記の「ボジョレー」を含むウェブページだけでなく、日本語のカタカナ表記の「ボジョレー」と翻字関係にあるアルファベット表記の「Beaujolais」を含むウェブページも検索結果として提示することが可能となる。

[0004] 従来、このような翻字辞書データの作成（すなわち、翻字辞書データへの翻字対の登録）は、人による手作業や、予め用意された学習データに基づく機械学習等によって行われている（例えば下記特許文献１参照）。

先行技術文献

特許文献

[0005] 特許文献１：特開２００７－１５６５４５号公報

発明の概要

発明が解決しようとする課題

[0006] しかしながら、上述した翻字辞書データの作成を人が手作業で行う場合（すなわち、一の言語の単語が他の言語の単語と翻字関係にあるか否かを人が個々に判断する場合）には、非常に手間がかかる。一方、翻字辞書データの作成を機械学習により行う場合には、適切な翻字ルールを網羅的に生成するための学習データ（すなわち、互いに翻字関係にあることが判明している翻字対）を用意することが困難であるという問題がある。これは、適切な翻字ルールが作成するためにどのような学習ルールを用いて機械学習を行えばよいかを事前に把握することが困難であることに起因する。そこで、このような翻字辞書データの作成の効率化を図るために、翻字対を効率良く検出することが要請されている。

課題を解決するための手段

[0007] 本発明の一側面に係る翻字処理装置は、所定の文字体系を有する第1の言語で記述された第1の単語をアルファベット表記で表した第1のアルファベット文字列と、第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得部と、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて第1の単語と第2の単語とが翻字関係にあるか否かを決定する決定部と、決定部によって翻字関係にあると決定された第1の単語及び第2の単語を翻字対として出力する出力部と、を備える。

[0008] 本発明の一側面に係る翻字処理方法は、少なくとも一つのプロセッサにより実行される翻字処理方法であって、所定の文字体系を有する第1の言語で記述された第1の単語をアルファベット表記で表した第1のアルファベット文字列と、第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得ステップと、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素と

が予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて第1の単語と第2の単語とが翻字関係にあるか否かを決定する決定ステップと、決定ステップにおいて翻字関係にあると決定された第1の単語及び第2の単語を翻字対として出力する出力ステップと、を含む。

[0009] 本発明の一側面に係る翻字処理プログラムは、所定の文字体系を有する第1の言語で記述された第1の単語をアルファベット表記で表した第1のアルファベット文字列と、第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得ステップと、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて第1の単語と第2の単語とが翻字関係にあるか否かを決定する決定ステップと、決定ステップにおいて翻字関係にあると決定された第1の単語及び第2の単語を翻字対として出力する出力ステップと、をコンピュータに実行させる。

[0010] このような側面においては、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定結果に基づいて、第1の言語で記述された第1の単語と第2の言語で記述された第2の単語とが翻字関係にあるか否かが決定される。このように、互いに異なる文字体系を有する言語で記述された単語同士が翻字関係にあるか否かを、それぞれの単語をアルファベット表記で表したアルファベット文字列に含まれる子音要素間の対応関係に着目して決定することにより、翻字対を効率良く検出することができる。

発明の効果

[0011] 本発明の一側面によれば、翻字対を効率良く検出することができる。

図面の簡単な説明

[0012] [図1]実施形態における翻字処理装置を含む情報処理システムの機能構成を示す図である。

[図2]翻字辞書データの一例を示す図である。

[図3]実施形態における情報処理システムに用いられるコンピュータのハードウェア構成を示す図である。

[図4]E Cサイトのウェブページの構成の例を示す図である。

[図5]単語取得部、文字列取得部、及び分割部の処理の一例を示す図である。

[図6]単語取得部、文字列取得部、及び分割部の処理の一例を示す図である。

[図7]子音ルール及び禁止ルールの一例を示す図である。

[図8]実施形態に係る翻字処理装置の動作を示すフローチャートである。

[図9]図8におけるステップS4の処理の詳細を示すフローチャートである。

[図10]実施形態に係る翻字処理プログラムの構成を示す図である。

発明を実施するための形態

[0013] 以下、添付図面を参照しながら本発明の実施形態を詳細に説明する。なお、図面の説明において同一又は同等の要素には同一の符号を付し、重複する説明を省略する。

[0014] 図1～図7を用いて、実施形態に係る翻字処理装置10を含む情報処理システム1の機能及び構成を説明する。情報処理システム1は、翻字処理装置10と、情報処理装置20と、を含む。

[0015] 翻字処理装置10は、所定の文字体系を有する第1の言語で記述された第1の単語と、第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語とが翻字関係にあるか否かを決定し、翻字関係にあると決定された単語のペアを翻字対として出力する装置である。具体的には、翻字処理装置10は、第1の単語をアルファベット表記で表した第1のアルファベット文字列と、第2の単語をアルファベット表記で表した第2のアルファベット文字列とを取得する。そして、翻字処理装置10は、第1のアルファベット文字列に含まれる子音要素と第2のアルファベット文字列に含まれる子音要素とが予め定められた対応関係にあるか否かの判定の結果に基づいて、第1の単語と第2の単語とが翻字関係にあるか否かを決定する。

[0016] 「文字体系」とは、世界で使用されている書き言葉（文字）の分類区分を

意味する。文字体系の具体例としては、ラテン文字、アラビア文字、ギリシア文字、キリル文字、漢字、かな文字、ハングル等の分類区分が挙げられる。また、「ある言語 a がある文字体系 b を有する」とは、当該言語 a において使用される文字の分類区分が当該文字体系 b であることを意味する。例えば、ラテン文字が使用される英語、フランス語、及びドイツ語等は、文字体系としてラテン文字を有する。同様に、ロシア語は文字体系としてキリル文字を有し、中国語は文字体系として漢字を有し、日本語は文字体系として漢字及びかな文字（ひらがな及びカタカナ）を有する。

[0017] 「アルファベット表記」とは、個々の文字が子音又は母音の音素を表す音素文字による表記全般を意味する。本実施形態では、アルファベット表記の一例として、「A」から「Z」までの26個のラテン文字によるラテン文字表記を用いるが、アルファベット表記は、例えばギリシア文字やキリル文字等のラテン文字以外の文字による表記も含む概念であってよい。

[0018] 「ある言語 a の単語 w をアルファベット表記で表したアルファベット文字列 x」は、当該単語 w がアルファベット表記（本実施形態では上述の通りラテン表記を意味する。以下同様。）で記述されている場合（例えば言語 a が英語である場合等）には、当該単語 w と同一の文字列である。一方、当該単語 w がアルファベット表記で記述されていない場合（例えば言語 a が日本語であり、単語 w がカタカナ表記で記述されている場合等）には、上記アルファベット文字列 x は、当該単語 w をアルファベット表記（例えばローマ字表記等）に変換することにより得られる文字列である。

[0019] 「翻字」とは、一の文字体系で記述された文字を他の文字体系による文字に変換することを意味する。ここで、一の文字体系で記述された1文字が他の文字体系で記述された1以上の文字に変換されてもよいし、一の文字体系で記述された2以上の文字が他の文字体系で記述された1以上の文字に変換されてもよい。

[0020] 一の言語 L 1 の単語 W 1 と当該言語 L 1 とは異なる文字体系を有する他の言語 L 2 の単語 W 2 とが「翻字関係にある」とは、単語 W 2 が単語 W 1 を翻

字することにより得られる単語であること（すなわち、単語W1を言語L2による表記に変換した単語が単語W2であること）及びその逆が成立することを意味する。例えば、カタカナ表記の「ボジョレー」は、アルファベット表記の「Beaujolais」をカタカナ表記に変換した単語であるため、カタカナ表記の「ボジョレー」とアルファベット表記の「Beaujolais」とは、翻字関係にあるといえる。「翻字対」とは、互いに翻字関係にある単語の組を意味する。例えば上述したカタカナ表記の「ボジョレー」とアルファベット表記の「Beaujolais」との組は、翻字対である。

[0021] 本実施形態では一例として、翻字処理装置10は、日本語（第1の言語）で記述されたカタカナ表記の単語（第1の単語。以下「カタカナ単語」という。）と、英語（第2の言語）で記述されたアルファベット表記の単語（第2の単語。以下「英単語」という。）とが翻字関係にあるか否かを決定する。ただし、翻字処理装置10によって翻字関係にあるか否かを決定する対象となる言語の組み合わせは、上記に限られず、互いに異なる文字体系を有する言語同士の組み合わせであれば何でもよい。なお、本実施形態で例示する英単語には、英語以外の言語に由来する単語（例えば上述したフランス語に由来する「Beaujolais」等）の英語表記が含まれる。

[0022] 情報処理装置20は、翻字処理装置10から出力される翻字対が登録された翻字辞書データTを参照することにより予め定められた処理（詳しくは後述）を実行する装置である。図2に、翻字辞書データTの一例を示す。図2に示すように、翻字辞書データTは、互いに翻字関係にあるカタカナ単語及び英単語からなる翻字対を複数含むデータである。ここで、翻字辞書データTの実装方式は限定されない。例えば、翻字辞書データTは、翻字処理装置10から情報処理装置20に受け渡されるテキストファイルであってもよいし、任意のデータベース管理システムによって管理されるデータベースであってもよい。なお、図2における翻字辞書データTの左側の列にはカタカナ単語（カタカナ表記）に対応するローマ字表記を併記しているが、実際の処

理においては、ローマ字表記は、翻字辞書データ T に登録されていなくてもよい。

[0023] 情報処理装置 20 が翻字辞書データ T を参照することにより実行する処理及び用途は、特定の処理及び用途に限定されない。以下では、情報処理装置 20 がインターネット上で商品（サービスを含む）を販売するウェブサイト（EC サイト）を構成するコンピュータ装置として実装される場合を例に挙げて、情報処理装置 20 が実行する処理のいくつかの例について説明する。なお、EC サイトにより販売される「サービス」の具体例としては、旅行のツアー及び宿泊の予約、コンサートチケットの予約、並びにゴルフ場の予約等が挙げられる。

[0024] （情報処理装置による処理の第 1 の例）

情報処理装置 20 は、翻字辞書データ T を参照することにより、EC サイトを利用するユーザによって入力される検索キーワードに基づく商品検索を実行してもよい。具体例として、カタカナ表記の「ボジョレー」がユーザによって検索キーワードとして入力された場合を考える。ここで、EC サイトに出品している販売者によって、同一の商品 X を示すために用いる表記が異なる場合がある。言い換えれば、EC サイト上において、商品 X をカタカナ表記の「ボジョレー」で表したウェブページもあれば、商品 X をカタカナ表記の「ボジョレー」と翻字関係にあるアルファベット表記の「Beaujolais」で表したウェブページもある場合がある。ここで、「ウェブページ」とは、インターネット等の通信ネットワークを介してアクセス可能な場所に保存された文書データである。ウェブページは、例えば HTML、スタイルシート、及び画像データ等を含み、ウェブブラウザによって閲覧可能なように構成されている。

[0025] 情報処理装置 20 は、ユーザによって入力されたカタカナ表記の「ボジョレー」を第 1 の検索キーワードとして受け付けるとともに、翻字辞書データ T を参照することにより、カタカナ表記の「ボジョレー」と翻字関係にあるアルファベット表記の「Beaujolais」を第 2 の検索キーワードと

して取得する。そして、情報処理装置20は、第1の検索キーワード又は第2の検索キーワードを含む商品のウェブページを検索結果としてユーザに提示する。これにより、ECサイトを利用するユーザに対して、当該ユーザが所望する商品Xに関するより多くの情報を1回の検索操作（この例では、ユーザがカタカナ表記の「ボジョレー」を検索キーワードとして入力する操作）に対して提供することができる。その結果、ECサイトを利用するユーザの利便性を向上させることができる。また、ECサイトに商品を出品する店舗側にとっても、自身が販売する商品のウェブページが検索結果としてユーザに提示される機会が増えるため、自身が販売する商品の販売機会を増やすことができるというメリットがある。また、ユーザの検索回数を減らすことができるため、ユーザによって入力されるデータによる通信負荷や情報処理装置20での検索処理にかかる処理負荷等を抑制することもできる。

[0026] （情報処理装置による処理の第2の例）

情報処理装置20は、翻字辞書データTを参照することにより、ユーザによる文字入力を補完する処理を実行してもよい。例えば、ECサイト等の検索フォームにユーザがカタカナ表記の「ボジョレー」を入力した際に、その変換候補として、カタカナ表記の「ボジョレー」と翻字関係にあるアルファベット表記の「Beaujolais」をユーザに提示してもよい。これにより、ユーザは、変換候補として提示されたアルファベット表記の「Beaujolais」を選択することで、アルファベット表記の「Beaujolais」を検索キーワードとして入力することができる。従って、例えば、アルファベット表記の「Beaujolais」を含む商品のウェブページを検索するためにアルファベット表記の「Beaujolais」を検索キーワードとして入力する必要があり、且つ、ユーザがアルファベット表記の「Beaujolais」のスペルを知らない場合であっても、ユーザは所望する検索を容易に実行することが可能となる。その結果、ECサイトを利用するユーザの利便性を向上させることができる。また、誤ったスペルの検索キーワードがユーザによって入力されることを抑制することができるた

め、無駄な検索処理の発生を抑制することもできる。

[0027] (情報処理装置による処理の第3の例)

情報処理装置20は、翻字辞書データTを参照することにより、商品の同一性を判定する処理を実行してもよい。例えば同一のECサイトにおいて多数の出品者が同一の商品Xを個別に販売するような場合において、同一の商品Xが販売者によって様々な表記で表される場合がある。例えば同一の商品Xについて、ある販売者はカタカナ表記の「ボジョレー」を使用し、他の販売者はアルファベット表記の「Beaujolais」を使用する場合があります。このような場合に、情報処理装置20は、翻字辞書データTを参照することにより、カタカナ表記の「ボジョレー」とアルファベット表記の「Beaujolais」とが互いに翻字関係にあることを検出し、カタカナ表記の「ボジョレー」とアルファベット表記の「Beaujolais」とが同一の商品を指していることを把握することができる。その結果、例えば、情報処理装置20は、カタカナ表記の「ボジョレー」が使用されている商品の販売数とアルファベット表記の「Beaujolais」が使用されている商品の販売数とを集計することにより、ECサイトでの商品Xの販売総数を正確に算出することができる。

[0028] 上記第1～第3の例で述べた通り、情報処理装置20は、翻字辞書データTを参照することにより、様々な目的に応じた処理を実行することができる。ただし、上記の第1～第3の例は翻字辞書データTを用いた処理の一例に過ぎない。情報処理装置20は、翻字辞書データTを参照することにより、第1～第3の例以外の処理を実行してもよい。

[0029] 図3に、翻字処理装置10又は情報処理装置20として機能するコンピュータ100の一般的なハードウェア構成を示す。コンピュータ100は、オペレーティングシステムやアプリケーション・プログラム等を実行するプロセッサ(例えばCPU(Central Processing Unit))101と、ROM(Read Only Memory)及びRAM(Random Access Memory)で構成される主記憶部102と、ハードディスクやフラッシュメモリ等で構成される補助

記憶部 103 と、ネットワークカード或いは無線通信モジュールで構成される通信制御部 104 と、キーボードやマウス等の入力装置 105 と、モニタ等の出力装置 106 と、を備える。当然ながら、搭載されるハードウェアモジュールはコンピュータ 100 の種類により異なる。

[0030] 情報処理システム 1 内の各機能要素は、プロセッサ 101 又は主記憶部 102 の上に所定のソフトウェアを読み込ませ、プロセッサ 101 の制御の下で通信制御部 104 や入力装置 105、出力装置 106 等を動作させ、主記憶部 102 又は補助記憶部 103 におけるデータの読み出し及び書き込みを行うことで実現される。処理に必要なデータやデータベースは、主記憶部 102 又は補助記憶部 103 内に格納される。

[0031] 翻字処理装置 10 及び情報処理装置 20 のそれぞれは、単一のコンピュータ 100 によって構成されてもよいし、複数のコンピュータ 100 間で処理を分散して実行する構成とされてもよい。また、翻字処理装置 10 及び情報処理装置 20 は、互いに異なるコンピュータ 100 によって構成されてもよいし、単一のコンピュータ 100 によって構成されてもよい。

[0032] 図 1 に示すように、翻字処理装置 10 は、機能的構成要素として、単語取得部 11 と、文字列取得部 12 と、分割部 13 と、決定部 14 と、ルール情報記憶部 15 と、出力部 16 と、を備える。翻字処理装置 10 の各構成要素について以下に説明する。

[0033] 単語取得部 11 は、所定の文字体系を有する第 1 の言語で記述された第 1 の単語と、第 1 の言語とは異なる文字体系を有する第 2 の言語で記述された第 2 の単語と、を取得する機能要素である。上述の通り、本実施形態では一例として、第 1 の言語は文字体系としてかな文字を有する日本語であり、第 1 の単語はカタカナ表記のカタカナ単語である。また、第 2 の言語は文字体系としてラテン文字を有する英語であり、第 2 の単語はアルファベット表記の英単語である。従って、単語取得部 11 は、カタカナ単語 w1 及び英単語 w2 を取得する。単語取得部 11 により取得されるカタカナ単語 w1 及び英単語 w2 は、文字列取得部 12 に受け渡され、後述の文字列取得部 12、分

割部 1 3、及び決定部 1 4 による処理を経て、翻字関係にあるか否かが決定される。

[0034] 単語取得部 1 1 は、カタカナ単語 w 1 及び英単語 w 2 の一方又は両方を複数取得してもよい。この場合、翻字処理装置 1 0 は、カタカナ単語 w 1 と英単語 w 2 とのすべての組み合わせについて、後述する文字列取得部 1 2、分割部 1 3、及び決定部 1 4 による処理を実行すればよい。このようにすれば、単語取得部 1 1 によって取得されたカタカナ単語 w 1 及び英単語 w 2 の複数の組み合わせのうちから、翻字関係にある組み合わせを自動的に検出することが可能となる。

[0035] 単語取得部 1 1 は、例えば、インターネット上に公開されている任意のウェブページに含まれるテキスト情報を入力として受け付け、当該テキスト情報に含まれるカタカナ表記の単語をカタカナ単語 w 1 として取得し、当該テキスト情報に含まれるアルファベット表記の単語を英単語 w 2 として取得してもよい。

[0036] 同一のウェブページに含まれるテキスト情報には、同一の概念（例えば、外来語で表される物の名称、外国の地名、外国人の氏名等）についてカタカナ表記及びアルファベット表記のそれぞれで表したカタカナ単語及び英単語が含まれている可能性が高い。言い換えれば、同一のウェブページに含まれるテキスト情報には、互いに翻字関係にあるカタカナ単語及び英単語が含まれている可能性が高い。このため、同一のウェブページに含まれるテキスト情報から翻字関係にあるか否かを決定する対象となるカタカナ単語 w 1 及び英単語 w 2 を取得することにより、翻字対を効率良く検出することが期待できる。

[0037] また、図 4 に示すように、EC サイトにおいて商品又はサービスを提示するウェブページ（以下「商品ページ」と言う。）には、通常、商品の写真、価格、概要、詳細、レビューや商品を購入候補として選択する選択ボタン等の専用領域が設けられるとともに、商品又はサービスの名称を記載するための専用領域（商品名領域 R）が設けられる。

[0038] ここで、商品又はサービスの名称が、カタカナ表記で表される場合（例えば、商品又はサービスの名称が、外来語で表される物の名称、外国の地名、外国人の氏名等に基づくものである場合）には、商品又はサービスの名称のカタカナ表記及びアルファベット表記の両方が商品名領域 R に含まれている可能性が高い。これは、EC サイトにおける商品ページは、なるべく多くのユーザの目に触れられることが好ましいからである。すなわち、商品ページの管理者（例えば当該商品ページに提示される商品又はサービスの販売者）は、インターネット上でのウェブページ検索において、当該商品ページがなるべく多くの検索キーワードにより検索ヒットすることを望む。このような事情から、商品名領域 R には、商品又はサービスの名称のカタカナ表記及びアルファベット表記の両方が記載されている可能性が高い。

[0039] そこで、単語取得部 11 は、EC サイトの商品ページにおける商品名領域 R（商品又はサービスの名称の記載箇所）を特定し、当該商品名領域 R に含まれるカタカナ表記の単語をカタカナ単語 w1 として取得し、当該商品名領域 R に含まれるアルファベット表記の単語を英単語 w2 として取得してもよい。これにより、翻字関係にある可能性の高いカタカナ単語 w1 及び英単語 w2 を抽出することができ、翻字対をより効率良く検出することが期待できる。ここで、単語取得部 11 が商品名領域 R を特定する方法は何でもよいが、単語取得部 11 は、例えば商品ページのレイアウトを示すレイアウト情報（例えば HTML 等で記述された情報）を参照することにより、商品名領域 R を特定してもよい。

[0040] ただし、単語取得部 11 がカタカナ単語 w1 及び英単語 w2 を取得する手法は上記の手法に限られない。例えば、単語取得部 11 は、翻字処理装置 10 のオペレータによって予め準備されたカタカナ単語 w1 及び英単語 w2 の組（或いは集合）を入力として受け付けてもよい。

[0041] 文字列取得部 12 は、第 1 の単語をアルファベット表記で表した第 1 のアルファベット文字列と、第 2 の単語をアルファベット表記で表した第 2 のアルファベット文字列と、を取得する機能要素である。具体的には、文字列取

得部 1 2 は、単語取得部 1 1 により取得された第 1 の単語及び第 2 の単語（本実施形態では一例として、カタカナ単語 w 1 及び英単語 w 2）を必要に応じてアルファベット文字列に変換する機能要素である。

[0042] 図 5 に示すように、カタカナ単語 w 1 がカタカナ表記の「ボジョレー」であり、英単語 w 2 がアルファベット表記の「BEAUJOLAIS」である場合を例として、文字列取得部 1 2 が実行する処理について説明する。なお、本明細書の以降の説明においては、英単語についてはすべて大文字で表すが、英単語 w 2 が小文字を含んでいてもよいことは言うまでもない。なお、英単語 w 2 が小文字を含んでいる場合には、文字列取得部 1 2 は、大文字のみを用いることで文字列処理の単純化を図るために、英単語 w 2 に含まれるすべての小文字を大文字に変換してもよい。文字列取得部 1 2 は、例えば予め用意された大文字と小文字との対応を示す変換テーブルを参照することにより、英単語 w 2 に含まれる小文字を当該小文字に対応する大文字に変換することができる。また、文字列取得部 1 2 は、小文字のみを用いることで文字列処理の単純化を図る場合には、上述した手法と同様の手法により、英単語 w 2 に含まれるすべての大文字を小文字に変換してもよい。

[0043] 図 5 の（a）に示すように、文字列取得部 1 2 は、カタカナ単語 w 1 をローマ字表記（アルファベット表記）で表したローマ字文字列（第 1 のアルファベット文字列）s t r 1 に変換する。文字列取得部 1 2 は、例えば予め用意されたカタカナ文字とローマ字との対応を示す変換テーブルを参照することにより、カタカナ単語 w 1 に含まれるカタカナ文字を、当該カタカナ文字に対応するローマ字に変換することができる。図 5 の（a）の例では、文字列取得部 1 2 は、上述した変換テーブルを用いた変換処理を実行することにより、カタカナ表記の「ボ」、「ジョ」、「レ」、「（エ）ー」のそれぞれをローマ字表記の「BO」、「JO」、「RE」、「E」に変換し、「BOJOREE」を示すローマ字文字列 s t r 1 を取得する。

[0044] 一方、本実施形態では、図 5 の（b）に示すように、英単語 w 2 は既にアルファベット表記で表されている。従って、文字列取得部 1 2 は、単語取得

部 1 1 により取得された英単語 w_2 をそのまま英語文字列（第 2 のアルファベット文字列） $s t r_2$ として取得する。

[0045] 分割部 1 3 は、ローマ字文字列 $s t r_1$ からローマ字配列（第 1 の配列） A を生成し、英語文字列 $s t r_2$ から英語配列（第 2 の配列） B を生成する機能要素である。具体的には、分割部 1 3 は、ローマ字文字列 $s t r_1$ を母音字から子音字に切り替わる境界で分割し、分割された個々の要素をローマ字文字列 $s t r_1$ における出現順に配列したローマ字配列 A を生成する。また、分割部 1 3 は、英語文字列 $s t r_2$ を母音字から子音字に切り替わる境界で分割し、分割された個々の要素を英語文字列 $s t r_2$ における出現順に配列した英語配列 B を生成する。

[0046] ローマ字文字列 $s t r_1$ 又は英語文字列 $s t r_2$ に、母音字から子音字に切り替わる境界がない場合には、分割部 1 3 は元のローマ字文字列 $s t r_1$ 又は英語文字列 $s t r_2$ を単一の要素とするローマ字配列 A 又は英語配列 B を生成すればよい。母音字から子音字に切り替わる境界がない文字列の例としては、「 $d e$ 」等の子音要素と母音要素とがそれぞれ一つしかない文字列等がある。

[0047] 「母音字」とは、アルファベット表記において母音を表す文字を意味する。上述の通り、本実施形態では一例として、アルファベット表記はラテン文字表記であるため、母音字は、「 A 」、「 I 」、「 U 」、「 E 」、「 O 」である。一方、「子音字」とは、アルファベット表記において子音を表す文字（すなわち母音字以外の文字）を意味する。すなわち、子音字は、「 A 」から「 Z 」までの 26 個の文字から上述した 5 つの母音字を除いた 21 個の文字である。

[0048] 以下、図 5 及び図 6 を用いて、分割部 1 3 によって生成されるローマ字配列 A 及び英語配列 B について説明する。なお、以下に説明する配列の表現形式は一例に過ぎず、分割部 1 3 によって生成される配列の具体的なデータ構造を限定するものではない。

[0049] （分割部による処理の第 1 の例）

図5を用いて、分割部13による処理の第1の例について説明する。図5の(a)に示すように、分割部13は、ローマ字文字列 `s t r 1` (“B O J O R E E”) を母音字から子音字に切り替わる境界で分割することにより、「B O」、「J O」、「R E E」の3つの要素に分割する。

[0050] このような分割処理は、例えば以下のようにして行われる。すなわち、分割部13は、予め母音字である文字の情報（以下「母音字情報」）を記憶しておく。そして、分割部13は、処理対象のローマ字文字列 `s t r 1` を先頭の文字から順に1文字ずつ読み込み、母音字情報を参照することにより、読み込まれた文字が母音字であるか子音字であるかを判定する。そして、分割部13は、母音字であると判定された文字の次の文字が子音字であると判定された場合に、母音字であると判定された文字と子音字であると判定された文字との間を境界として、ローマ字文字列 `s t r 1` を分割する。

[0051] 分割部13は、分割された個々の要素を更に母音要素と子音要素とに区別した上で、ローマ字文字列 `s t r 1` における出現順に配列することにより、ローマ字配列A(= { [“B” “O”], [“J” “O”], [“R” “E E”] }) を生成する。ここで、「子音要素」は一以上の子音字からなる文字列を意味し、「母音要素」は一以上の母音字からなる文字列を意味する。

[0052] 一方、図5の(b)に示すように、分割部13は、英語文字列 `s t r 2` (“B E A U J O L A I S”) についても、上述のローマ字文字列 `s t r 1` に対する処理と同様に、母音字から子音字に切り替わる境界で分割することにより、「B E A U」、「J O」、「L A I」、「S」の4つの要素に分割する。分割部13は、分割された個々の要素を更に母音要素と子音要素とに区別した上で、英語文字列 `s t r 2` における出現順に配列することにより、英語配列B(= { [“B” “E A U”], [“J” “O”], [“L” “A I”], [“S” “”] }) を生成する。

[0053] ここで、本明細書で用いる配列の表記方法について説明する。図5に示すように、本明細書では、配列の個々の要素を括弧[]で示し、括弧{}内に配列を構成する各要素をカンマ区切りで並べることで配列を表現する。括弧

[] 内の左側部分は当該要素に含まれる子音要素を示し、括弧 [] 内の右側部分は当該要素に含まれる母音要素を示す。母音要素がない要素については、括弧 [] 内の右側部分は空文字 (“ ”) となり、子音要素がない要素については、括弧 [] 内の左側部分は空文字 (“ ”) となる。また、以降の説明において配列 X の i 番目の要素を X [i] と表記する。

[0054] (分割部による処理の第2の例)

次に、図6を用いて、分割部13による処理の第2の例について説明する。図6の例では、単語取得部11により取得されるカタカナ単語 w 1 はカタカナ表記の「バスタオル」であり、単語取得部11により取得される英単語 w 2 はアルファベット表記の「BATH TOWEL」である。この場合、文字列取得部12は、カタカナ単語 w 1 からローマ字文字列 s t r 1 (“ B A S U T A O R U ”) を取得し、英単語 w 2 から英語文字列 s t r 2 (“ B A T H T O W E L ”) を取得する。

[0055] 図6の(a)に示すように、分割部13は、ローマ字文字列 s t r 1 (“ B A S U T A O R U ”) を母音字から子音字に切り替わる境界で分割することにより、「BA」、「SU」、「TAO」、「RU」の4つの要素に分割する。分割部13は、分割された個々の要素を更に母音字と子音字とに区別した上で、ローマ字文字列 s t r 1 における出現順に配列することにより、ローマ字配列 A (= { [“ B ” “ A ”] , [“ S ” “ U ”] , [“ T ” “ A O ”] , [“ R ” “ U ”] }) を生成する。

[0056] 一方、図6の(b)に示すように、分割部13は、英語文字列 s t r 2 (“ B A T H T O W E L ”) を母音字から子音字に切り替わる境界で分割することにより、「BA」、「THTO」、「WE」、「L」の4つの要素に分割する。分割部13は、分割された個々の要素を更に母音字と子音字とに区別した上で、英語文字列 s t r 2 における出現順に配列することにより、英語配列 B 1 (= { [“ B ” “ A ”] , [“ T H T ” “ O ”] , [“ W ” “ E ”] , [“ L ” “ ”] }) を生成する。

[0057] ここで、英語配列 B 1 の2番目の要素 B 1 [2] (= [“ T H T ” “ O ”

〕)には、複数の子音字（この例では、“T”、“H”、“T”の3つの子音字）からなる子音要素（“THT”）が含まれている。この場合、分割部13は、当該要素B1〔2〕を二以上の要素に更に分割する分割パターンに対応する配列B2～B4を生成する。具体的には、分割部13は、要素B1〔2〕に含まれる子音要素“THT”について可能なすべての分割パターンに対応する英語配列B2～B4を生成する。

[0058] 図6の(b)に示すように、英語配列B2（＝{〔“B” “A”〕, 〔“TH” “”〕, 〔“T” “O”〕, 〔“W” “E”〕, 〔“L” “”〕}）は、英語配列B1の要素B1〔2〕に含まれる子音要素“THT”を“TH”と“T”とに分割する分割パターンに対応する配列である。すなわち、英語配列B2は、英語配列B1の要素B1〔2〕を要素B2〔2〕（＝〔“TH” “”〕）と要素B3〔2〕（＝〔“T” “O”〕）とに分割することにより得られる配列である。

[0059] 英語配列B3（＝{〔“B” “A”〕, 〔“T” “”〕, 〔“HT” “O”〕, 〔“W” “E”〕, 〔“L” “”〕}）は、英語配列B1の要素B1〔2〕に含まれる子音要素“THT”を“T”と“HT”とに分割する分割パターンに対応する配列である。すなわち、英語配列B3は、英語配列B1の要素B1〔2〕を要素B3〔2〕（＝〔“T” “”〕）と要素B3〔3〕（＝〔“HT” “O”〕）とに分割することにより得られる配列である。

[0060] 英語配列B4（＝{〔“B” “A”〕, 〔“T” “”〕, 〔“H” “”〕, 〔“T” “O”〕, 〔“W” “E”〕, 〔“L” “”〕}）は、英語配列B1のB1〔2〕に含まれる子音要素“THT”を“T”と“H”と“T”とに分割する分割パターンに対応する配列である。すなわち、英語配列B4は、英語配列B1の要素B1〔2〕を要素B4〔2〕（＝〔“T” “”〕）と要素B4〔3〕（＝〔“H” “”〕）と要素B4〔4〕（＝〔“T” “O”〕）に分割することにより得られる配列である。

[0061] 決定部14は、カタカナ文字列s t r 1に含まれる子音要素（第1の子音要素）と英語文字列s t r 2に含まれる子音要素（第2の子音要素）とが予

め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいてカタカナ単語 w 1 と英単語 w 2 とが翻字関係にあるか否かを決定する機能要素である。また、ルール情報記憶部 15 は、日本語のローマ字表記において用いられる子音要素と、英語のアルファベット表記において用いられる子音要素との対応関係を示す子音ルールを記憶する機能要素である。決定部 14 による上記判定は、ルール情報記憶部 15 に記憶された子音ルールを参照することにより行われる。

[0062] 図 7 の (a) に、子音ルールの一例を示す。図 7 の (a) に示す子音ルールは、同じ行における左側のローマ字表記の子音要素と右側の英語表記の子音要素とが対応関係にあることを示す。この子音ルールは、例えばローマ字表記の子音要素 (“B” 、 “B Y”) と英語表記の子音要素 (“V” 、 “V V” 、 “W” 、 “B” 、 “B B” 、 “B H” 、 “B J” 、 “B Y”) とが対応関係にあること等を示す。

[0063] ここで、「ローマ字表記の子音要素 c 1 と英語表記の子音要素 c 2 とが対応関係にある」とは、同一の概念を表すローマ字表記及び英語表記 (すなわち、互いに翻字関係にあるローマ字表記及び英語表記) の組において、ローマ字表記の子音要素 c 1 が英語表記においては子音要素 c 2 で表される場合があること (或いは、英語表記の子音要素 c 2 がローマ字表記においては子音要素 c 1 で表される場合があること) を意味する。

[0064] このような子音ルールは、例えば以下のような処理をオペレータ等が予め手作業等によって実行することで用意される。すなわち、いくつかの英単語について、当該英単語の発音をカタカナ表記で表し、更に当該カタカナ表記をローマ字表記に変換する処理を実行する。そして、当該英単語の子音要素と当該ローマ字表記の子音要素とを同一の発音箇所に対応する部分同士で比較することにより、互いに対応関係にある子音要素の組を抽出する。一例として、英単語の「V I B R A T I O N」と、当該英単語に対応する日本語のローマ字表記の「B A I B U R E E S H O N」とから互いに対応関係にある子音要素の組を抽出する場合について説明する。

[0065] この場合、ローマ字表記の「BAIBUREESHON」の「BA」及び「BU」の発音箇所に対応する部分同士と比較により、ローマ字表記の子音要素「B」と英語表記の子音要素「B」、「V」とが互いに対応関係にある子音要素の組として抽出される。また、ローマ字表記の「RE」の発音箇所に対応する部分同士と比較により、ローマ字表記の子音要素「R」と英語表記の子音要素「R」とが互いに対応関係にある子音要素の組として抽出される。また、ローマ字表記の「SHO」の発音箇所に対応する部分同士と比較により、ローマ字表記の子音要素「SH」と英語表記の子音要素「T」とが互いに対応関係にある子音要素の組として抽出される。このような抽出処理を複数の英単語とそのローマ字表記とについて行うことによって、図7の（a）に示すような子音ルールが作成される。

[0066] ここで、互いに翻字関係にあるカタカナ単語及び英単語における母音要素間の対応関係は複雑でありルール化することが困難であるのに対し、子音要素間の対応関係はある程度限定された数のルールにより規定されるということが、本発明者によって見出された。このような知見に基づき、決定部14が、図7の（a）に示すような予め用意された子音ルールを参照して子音要素の対応関係を判定することにより、以下に例示する決定処理が実現される。以下に、決定部14により実行される処理の第1～第4の例について説明する。

[0067] （決定部による処理の第1の例）

第1の例では、決定部14は、ローマ字配列Aの各要素と英語配列Bの各要素とを先頭の要素から順に要素ごとに比較する。そして、決定部14は、ローマ字配列Aの各要素に含まれる子音要素と英語配列Bの各要素に含まれる子音要素とが対応関係にあると判定された場合に、カタカナ単語w1と英単語w2とが翻字関係にあると決定する。

[0068] 単語取得部11により取得されるカタカナ単語w1及び英単語w2、文字列取得部12により取得されるローマ字文字列str1及び英語文字列str2、並びに分割部13により生成されるローマ字配列A及び英語配列Bが

以下に示す場合を例として、決定部 14 により実行される処理の第 1 の例について詳細に説明する。

w 1 = カタカナ表記の「プリムール」

w 2 = 「P R I M E U R」

s t r 1 = 「P U R I M U U R U」

s t r 2 = 「P R I M E U R」

A = { [“P” “U”], [“R” “I”], [“M” “U U”], [“R” “U”] }

B = { [“P” “”], [“R” “I”], [“M” “E U”], [“R” “”] }

[0069] また、ルール情報記憶部 15 に記憶された子音ルールには、ローマ字表記の子音要素 “P” 及び英語表記の子音要素 “P” の組、ローマ字表記の子音要素 “R” 及び英語表記の子音要素 “R” の組、並びにローマ字表記の子音要素 “M” 及び英語表記の子音要素 “M” の組がそれぞれ対応関係にあることを示すルールが少なくとも含まれているものとする。

[0070] この場合、決定部 14 は、まず、ローマ字配列 A の 1 番目の要素 A [1] の子音要素 “P” と英語配列 B の 1 番目の要素 B [1] の子音要素 “P” とを比較する。決定部 14 は、子音ルールを参照することにより、これらの子音要素が互いに対応関係にあると判定する。続いて、決定部 14 は、ローマ字配列 A の 2 番目の要素 A [2] の子音要素 “R” と英語配列 B の 2 番目の要素 B [2] の子音要素 “R” とを比較する。決定部 14 は、子音ルールを参照することにより、これらの子音要素が互いに対応関係にあると判定する。

[0071] 続いて、決定部 14 は、ローマ字配列 A 及び英語配列 B のすべての要素についての比較を完了するまで、残りの要素についても同様の処理を順に実行する。これにより、決定部 14 は、ローマ字配列 A の 3 番目の要素 A [3] の子音要素 “M” と英語配列 B の 3 番目の要素 B [3] の子音要素 “M” とが互いに対応関係にあると判定し、ローマ字配列 A の 4 番目の要素 A [4]

の子音要素“R”と英語配列Bの4番目の要素B〔4〕の子音要素“R”とが互いに対応関係にあると判定する。

[0072] このようにローマ字配列Aの各要素に含まれる子音要素と英語配列Bの各要素に含まれる子音要素とが対応関係にあると判定された場合には、決定部14は、カタカナ単語w1と英単語w2とが翻字関係にあると決定する。

[0073] 一方、ローマ字配列Aの子音要素と英語配列Bの子音要素との要素ごとの比較のいずれかにおいて、子音要素同士が対応関係にあると判定されなかった場合には、決定部14は、カタカナ単語w1と英単語w2とが翻字関係にないと決定する。例えば、ローマ字配列Aの2番目の要素A〔2〕の子音要素と英語配列Bの2番目の要素A〔2〕の子音要素との組が、互いに対応関係にある子音要素の組として子音ルールに含まれていない場合（例えば要素A〔2〕の子音要素が“B”であり、要素B〔2〕の子音要素が“T”である場合等）、決定部14は、要素A〔2〕の子音要素と要素B〔2〕の子音要素とが対応関係にないと判定する。その結果、決定部14は、カタカナ単語w1と英単語w2とが翻字関係にないと決定する。

[0074] ここで、上述した理由により、ローマ字表記のアルファベット文字列に含まれる各子音要素と英語表記のアルファベット文字列に含まれる各子音要素とが対応関係にある場合には、カタカナ単語w1と英単語w2とが翻字関係にある可能性が高いといえる。従って、上述した決定部14の処理により、カタカナ単語w1と英単語w2とが翻字関係にあるか否かを子音要素同士の対応関係に基づいて容易に切り分けることができ、翻字対を効率良く検出することができる。

[0075] 以下、上述した子音要素同士の対応関係に基づき翻字対を決定する手法の効果について、機械学習を行う場合と対比して説明する。例えば機械学習により翻字辞書データを作成する場合には、適切な翻字ルール（文字間の変換ルール）を生成するために様々な学習データによる機械学習を試行錯誤する必要がある。これは、どのような学習データを用いて機械学習をすればどのような翻字ルールが生成されるかを事前に把握することが困難であることや

、単純に学習データを多くすれば精度の良い翻字ルールが生成されるわけではないこと（過学習により不適切な翻字ルールが生成されるおそれがあること）等に起因する。一方、上述した子音要素同士の対応関係に基づき翻字対を決定する手法によれば、予め定められた子音ルールに基づく判定によって比較対象となる２つの単語（カタカナ単語 w_1 及び英単語 w_2 ）が翻字関係にあるか否かを切り分けることができる。このため、上述したような機械学習における手間（学習データの選定や機械学習により得られた翻字ルールの適否の検証等）がかからない。

[0076] （決定部による処理の第２の例）

第２の例では、決定部 14 は、ローマ字文字列 $s t r 1$ に含まれる子音要素及び英語文字列 $s t r 2$ に含まれる子音要素のうち、予め定めた条件に合致する子音要素を除外し、残される子音要素について上記判定を実行する。

[0077] カタカナ単語 w_1 と英単語 w_2 とが翻字関係にある場合において、一方の単語のアルファベット文字列に現れる子音要素に対応する子音要素が、他方の単語のアルファベット文字列には明示的に現れない場合がある。例えば、英語文字列 $s t r 2$ において特定の態様で現れる子音要素に対応する子音要素は、ローマ字文字列 $s t r 1$ には現れないことがある。そこで、決定部 14 は、このような特定の態様で出現する子音要素を予め定めた条件に合致するか否かの判定によって抽出し、抽出された子音要素を除外することによって、カタカナ単語 w_1 と英単語 w_2 とが翻字関係にあるか否かをより精度良く決定することが可能となる。

[0078] 以下、図 5 に示す場合（カタカナ単語 w_1 がカタカナ表記の「ボジョレー」であり、英単語 w_2 がアルファベット表記の「BEAUJOLAIS」である場合）を例として、決定部 14 により実行される処理の第２の例について詳細に説明する。この場合において、ルール情報記憶部 15 に記憶された子音ルールには、ローマ字表記の子音要素“B”及び英語表記の子音要素“B”の組、ローマ字表記の子音要素“J”及び英語表記の子音要素“J”の組、並びにローマ字表記の子音要素“R”及び英語表記の子音要素“L”の

組がそれぞれ対応関係にあることを示すルールが少なくとも含まれているものとする。

[0079] また、ルール情報記憶部 15 は、子音ルールとともに、特定の子音要素の除外可能な条件を示す除外ルールを記憶する。ここでは一例として、ルール情報記憶部 15 は、英語文字列 `s t r 2` に現れる子音要素 “S” を除外可能な条件（出現位置が文字列の最後尾であること）を示す除外ルールを記憶しているものとする。

[0080] この場合、決定部 14 は、第 1 の例で述べた手順と同様の手順により、先頭の要素から順にローマ字配列 A の要素と英語配列 B の要素とを比較する。これにより、決定部 14 は、ローマ字配列 A の 1 番目の要素 A [1] の子音要素 “B” と英語配列 B の 1 番目の要素 B [1] の子音要素 “B” とが互いに対応関係にあり、ローマ字配列 A の 2 番目の要素 A [2] の子音要素 “J” と英語配列 B の 2 番目の要素 B [2] の子音要素 “J” とが互いに対応関係にあり、ローマ字配列 A の 3 番目の要素 A [3] の子音要素 “R” と英語配列 B の 3 番目の要素 B [3] の子音要素 “L” とが互いに対応関係にあると判定する。

[0081] このようにローマ字配列 A の 3 番目の要素 A [3] と英語配列 B の 3 番目の要素 B [3] との比較を完了した時点で、ローマ字配列 A のすべての要素についての比較が完了する。一方、英語配列 B の 4 番目の要素 B [4] (= [“S” “”]) が残る。この場合、決定部 14 は、当該要素 B [4] の子音要素を比較処理の対象から除外可能か否かを、ルール情報記憶部 15 に記憶された除外ルールを参照することにより判定する。

[0082] 上述の通り、除外ルールには、英語文字列 `s t r 2` の子音要素 “S” の出現位置が文字列の最後尾である場合に除外可能であることを示すルールが含まれている。また、要素 B [4] は、子音要素が “S” であり、出現位置が英語文字列 `s t r 2` の最後尾である（すなわち、英語配列 B の最後の要素である）。従って、決定部 14 は、要素 B [4] の子音要素が除外ルールに示される条件に合致すると判定し、当該要素 B [4] を比較の対象から除外す

る。その結果、比較の対象から除外された要素B〔4〕以外のすべての要素について、ローマ字配列Aの各要素に含まれる子音要素と英語配列Bの各要素に含まれる子音要素とが対応関係にあると判定されることになる。これにより、決定部14は、カタカナ単語w1と英単語w2とが翻字関係にあると決定する。

[0083] なお、第2の例では、除外可能な子音要素が英語配列Bの最後の要素B〔4〕に現れる場合を例として説明したが、除外可能な子音要素は、英語配列Bの最後の要素以外の要素に現れてもよい。除外可能な子音要素が英語配列Bの途中の要素に現れる場合については、後述する第3の例の中で説明する。

[0084] （決定部による処理の第3の例）

第3の例は、分割部13が複数の分割パターンの各々に対応するローマ字配列及び英語配列の組を生成する場合に対応する。すなわち、第3の例では、決定部14は、複数の分割パターンの各々について、ローマ字配列の各要素と英語配列の各要素とを先頭の要素から順に要素ごとに比較する。そして、決定部14は、少なくとも一つの分割パターンについてローマ字配列の各要素に含まれる子音要素と英語配列の各要素に含まれる子音要素とが対応関係にあると判定された場合に、カタカナ単語w1と英単語w2とが翻字関係にあると決定する。

[0085] 以下、図6に示す場合（カタカナ単語w1がカタカナ表記の「バスタオル」であり、英単語w2が「BATH TOWEL」である場合）を例として、決定部14により実行される処理の第3の例について詳細に説明する。

[0086] この場合において、ルール情報記憶部15に記憶された子音ルールには、ローマ字表記の子音要素“B”及び英語表記の子音要素“B”の組、ローマ字表記の子音要素“S”及び英語表記の子音要素“TH”の組、ローマ字表記の子音要素“T”及び英語表記の子音要素“T”の組、並びにローマ字表記の子音要素“R”及び英語表記の子音要素“L”の組がそれぞれ対応関係にあることを示すルールが少なくとも含まれているものとする。一方、子音

ルールには、ローマ字表記の子音要素“S”と英語表記の子音要素“T H T”又は“T”との組が互いに対応関係にあることを示すルールは含まれていないものとする。

[0087] また、ルール情報記憶部15は、英語文字列s t r 2に現れる子音要素“W”を除外可能な条件（子音字“W”のみを含む子音要素であること）を示す除外ルールを記憶しているものとする。一方、ルール情報記憶部15には、英語文字列s t r 2に現れる子音要素“T H T”、“H T”、“T”を除外可能な条件を示す除外ルールは記憶されていないものとする。

[0088] 図6の例では、「分割部による処理の第2の例」において述べた通り、分割部13により、英単語w 2から得られる英語文字列s t r 2（“B A T H T O W E L”）の4つの分割パターンの各々について、英語配列B 1～B 4が生成される。従って、決定部14は、ローマ字配列Aと英語配列B 1～B 4とのすべての組み合わせ（すなわち、ローマ字配列A及び英語配列B 1のペア、ローマ字配列A及び英語配列B 2のペア、ローマ字配列A及び英語配列B 3のペア、並びにローマ字配列A及び英語配列B 4のペア）について、第1の例及び第2の例で述べた処理と同様の処理を実行する。

[0089] まず、決定部14は、ローマ字配列A及び英語配列B 1のペアについて、ローマ字配列Aの各要素の子音要素と英語配列B 1の各要素の子音要素とが対応関係にあるか否かを判定する。この場合、決定部14は、1番目の要素同士の比較において、要素A [1]の子音要素“B”と要素B 1 [1]の子音要素“B”とが対応関係にあると判定する。

[0090] 一方、2番目の要素同士の比較においては、決定部14は、要素A [2]の子音要素“S”と要素B 1 [2]の子音要素“T H T”とが対応関係にないと判定する。そこで、決定部14は、ルール情報記憶部15に記憶された除外ルールを参照することにより、子音要素“T H T”を除外可能であるか否かを判定する。ここで、子音要素“T H T”を除外可能な条件を示すルールは除外ルールに含まれていない。従って、決定部14は、要素B 1 [2]を比較対象から除外できないと判定し、ローマ字配列A及び英語配列B 1の

ペアについては、ローマ字配列 A の各要素に含まれる子音要素と英語配列 B 1 の各要素に含まれる子音要素とが対応関係にないと判定する。

[0091] 次に、決定部 14 は、ローマ字配列 A 及び英語配列 B 2 のペアについて、ローマ字配列 A の各要素の子音要素と英語配列 B 2 の各要素の子音要素とが対応関係にあるか否かを判定する。この場合、決定部 14 は、1 番目の要素同士と比較において、要素 A [1] の子音要素 “B” と要素 B 2 [1] の子音要素 “B” とが対応関係にあると判定する。また、2 番目の要素同士と比較においても、決定部 14 は、要素 A [2] の子音要素 “S” と要素 B 2 [2] の子音要素 “TH” とが対応関係にあると判定する。また、3 番目の要素同士と比較においても、決定部 14 は、要素 A [3] の子音要素 “T” と要素 B 2 [3] の子音要素 “T” とが対応関係にあると判定する。

[0092] 続いて、4 番目の要素同士と比較において、決定部 14 は、要素 A [4] の子音要素 “R” と要素 B 2 [4] の子音要素 “W” とが対応関係にないと判定する。そこで、決定部 14 は、ルール情報記憶部 15 に記憶された除外ルールを参照することにより、子音要素 “W” を除外可能であるか否かを判定する。ここで、要素 B 2 [4] の子音要素 “W” は、除外ルールに示される除外可能な条件に合致する。従って、決定部 14 は、要素 B 2 [4] を比較対象から除外し、比較処理を継続する。

[0093] 具体的には、決定部 14 は、要素 B 2 [4] の次の要素 B 2 [5] を要素 A [4] と比較する対象として選択し、要素 A [4] と要素 B 2 [5] とを比較する。この比較において、決定部 14 は、要素 A [4] の子音要素 “R” と要素 B 2 [5] の子音要素 “L” とが対応関係にあると判定する。その結果、比較の対象から除外された要素 B 2 [4] 以外のすべての要素について、ローマ字配列 A の各要素に含まれる子音要素と英語配列 B 2 の各要素に含まれる子音要素とが対応関係にあると判定されることになる。これにより、決定部 14 は、ローマ字配列 A 及び英語配列 B 2 のペアについては、ローマ字配列 A の各要素に含まれる子音要素と英語配列 B 2 の各要素に含まれる子音要素とが対応関係にあると判定する。

[0094] そして、決定部 14 は、少なくとも一つの分割パターン（すなわち、ローマ字配列 A 及び英語配列 B 2 のペア）について、ローマ字配列 A の各要素に含まれる子音要素と英語配列 B の各要素に含まれる子音要素とが対応関係にあると判定したことをもって、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にあると決定する。

[0095] なお、決定部 14 は、残りのペア（すなわちローマ字配列 A と英語配列 B 3, B 4 との各ペア）についての比較処理を実行する必要はない。ただし、ローマ字配列 A と英語配列 B 3, B 4 との各ペアについての比較をローマ字配列 A 及び英語配列 B 2 のペアについての比較よりも前、或いは並列的に実行する場合には、ローマ字配列 A と英語配列 B 3, B 4 との各ペアについての比較が実行されることになる。

[0096] 上記例のように、英語表記において複数の子音字からなる一つの子音要素（上記例では“T H T”）は、日本語のローマ字表記において母音要素によって分断される二以上の子音要素（上記例では“S”及び“T”）に対応する場合がある。従って、上記例のように、複数の子音字からなる子音要素を更に分割するパターン（上記例では英語配列 B 2, B 3, B 4 を生成する 3 つの分割パターン）と分割しないパターン（上記例では英語配列 B 1 を生成する分割パターン）とを含む複数の分割パターンの各々についての判定を実行することにより、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にあるか否かを精度良く検出することが可能となる。

[0097] （決定部による処理の第 4 の例）

第 4 の例では、決定部 14 は、ローマ字文字列 s t r 1 に含まれる母音要素（第 1 の母音要素）と英語文字列 s t r 2 に含まれる母音要素（第 2 の母音要素）との組み合わせに基づく判定を更に実行する。そして、決定部 14 は、当該判定の結果にも基づいてカタカナ単語 w 1 と英単語 w 2 とが翻字関係にあるか否かを決定する。このように、上述の第 1 ～第 3 の例において述べた子音要素同士の比較に加えて、更に母音要素同士の比較にも基づくことで、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にあるか否かを精度良く決

定することが可能となる。

[0098] 以下、決定部 14 による第 4 の例の処理について詳細に説明する。決定部 14 は、ローマ字配列 A の各要素の母音要素と英語配列 B の各要素の母音要素との各ペアについて、予め定められた禁止ルール（母音ルール）に該当するか否かを判定する。決定部 14 は、このような判定を、例えば図 7 の（b）に示すような禁止ルールを参照することにより行うことができる。このような禁止ルールは、例えば上述したルール情報記憶部 15 に子音ルール及び除外ルールとともに記憶されてもよい。

[0099] 図 7 の（b）に示す禁止ルールは、カタカナ単語 w 1 のローマ字文字列 s t r 1 における母音要素 “U” に対応する英単語 w 2 の英語文字列 s t r 2 における母音要素が “A” 又は “I” （ただし、“IU” 又は “IEU” を除く）である場合に、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にないことを示すルールである。このような禁止ルールは、翻字関係にあるカタカナ単語 w 1 及び英単語 w 2 において通常現れない母音要素同士のペアを規定することにより作成することができる。

[0100] 具体例として、カタカナ単語 w 1 がカタカナ表記の「プリムール」であり、英単語 w 2 が「PRIMEUR」である場合における母音要素同士の比較について説明する。この場合、上述の通り、ローマ字配列 A 及び英語配列 B は、以下のように表される。

A = { [“P” “U”], [“R” “I”], [“M” “UU”], [“R” “U”] }

B = { [“P” “”], [“R” “I”], [“M” “EU”], [“R” “”] }

[0101] 決定部 14 は、ローマ字配列 A の 1 番目の要素 A [1] の母音要素 “U” と英語配列 B の 1 番目の要素 B [1] の母音要素 “” （空文字）との組が、禁止ルールに該当するか否かを判定する。同様に、決定部 14 は、要素 A [2] の母音要素 “I” 及び要素 B [2] の母音要素 “I” の組、要素 A [3] の母音要素 “UU” 及び要素 B [3] の母音要素 “EU” の組、並びに要

素A [4] の母音要素 “U” 及び要素B [2] の母音要素 “” の組について、禁止ルールに該当するか否かを判定する。

[0102] この例では、図7の(b)に示す禁止ルールに該当する母音要素同士の組はないため、決定部14は、ローマ字文字列s t r 1に含まれる母音要素と英語文字列s t r 2に含まれる母音要素との組み合わせが禁止ルールに該当しないと判定する。この場合、決定部14は、子音要素同士の比較による判定の結果に基づいて、カタカナ単語w 1と英単語w 2とが翻字関係にあると決定する。

[0103] 一方、母音要素同士の判定において、禁止ルールに該当する母音要素の組み合わせが存在する場合には、決定部14は、ローマ字文字列s t r 1に含まれる母音要素と英語文字列s t r 2に含まれる母音要素との組み合わせが禁止ルールに該当すると判定する。この場合、決定部14は、上述の子音要素同士の比較による判定の結果にかかわらず、カタカナ単語w 1と英単語w 2とが翻字関係にないと決定する。

[0104] このように、決定部14は、子音要素による判定と母音要素による判定とを組み合わせ、いずれの判定においてもカタカナ単語w 1と英単語w 2とが翻字関係にあると判断される場合に、カタカナ単語w 1と英単語w 2とが翻字関係にあると決定することができる。これにより、カタカナ単語w 1と英単語w 2とが翻字関係にあるか否かの決定を精度良く実行することが可能となる。

[0105] ここで、子音要素同士の比較によって対応関係にあると判定された配列（ローマ字配列又は英語配列）の要素数が多くなるほど、カタカナ単語w 1と英単語w 2とが翻字関係にある可能性が高くなる一方で、母音要素同士の比較において例外的に禁止ルールに該当する可能性は高くなる。従って、子音要素同士の比較によって対応関係にあると判定された要素数が多い状況においては、母音要素同士の判定で禁止ルールに該当したことをもってカタカナ単語w 1と英単語w 2とが翻字関係にないと決定してしまうと、実際には翻字関係にあるカタカナ単語w 1と英単語w 2とが翻字関係にないと誤った決

定を下してしまうリスクが高くなる。

[0106] 一方、子音要素同士の比較によって対応関係にあると判定された配列の要素数が少なくなるほど、カタカナ単語 w_1 と英単語 w_2 とが翻字関係にある可能性が低くなる一方で、母音要素同士の比較において例外的に禁止ルールに該当する可能性は低くなる。従って、子音要素同士の比較によって対応関係にあると判定された要素数が少ない状況においては、母音要素同士の判定で禁止ルールに該当したことをもってカタカナ単語 w_1 と英単語 w_2 とが翻字関係にないと決定することにより、実際には翻字関係にないカタカナ単語 w_1 と英単語 w_2 とが翻字関係であると誤った決定を下してしまう可能性を低減することが期待できる。

[0107] そこで、決定部 14 は、子音要素同士の比較によって対応関係にあると判定された配列の要素数が予め定めた閾値以下である場合に限り、上述の母音要素同士の判定を実行してもよい。これにより、実際には翻字関係にあるカタカナ単語 w_1 と英単語 w_2 とが翻字関係にないと誤った決定を下してしまうリスクを抑制しつつ、実際には翻字関係にないカタカナ単語 w_1 と英単語 w_2 とが翻字関係であると誤った決定を下してしまう可能性を低減することが期待できる。

[0108] 出力部 16 は、決定部 14 によって翻字関係にあると決定されたカタカナ単語 w_1 及び英単語 w_2 を翻字対として出力する機能要素である。出力部 16 によって出力される翻字対は、例えば図 2 に示した翻字辞書データ T の 1 レコード分のデータとして登録され、情報処理装置 20 によって参照される。

[0109] 次に、図 8 及び図 9 を用いて、翻字処理装置 10 の動作（本実施形態に係る翻字処理方法を含む）について説明する。

[0110] まず、単語取得部 11 がカタカナ単語 w_1 及び英単語 w_2 を取得する（ステップ S1、単語取得ステップ）。続いて、文字列取得部 12 がカタカナ単語 w_1 をローマ字表記に変換することによりローマ字文字列 str_1 を取得し、英単語 w_2 をそのまま英語文字列 str_2 として取得する（ステップ S

2、文字列取得ステップ)。

[0111] 続いて、分割部13がローマ字文字列s t r 1及び英語文字列s t r 2のそれぞれを母音字から子音字に切り替わる境界で分割することにより、ローマ字配列A及び英語配列Bを生成する(ステップS3、分割ステップ)。ここで、上述した「分割部による処理の第2の例」のように、英語文字列s t r 2が複数の子音字からなる子音要素を含む場合には、当該子音要素を更に分割する分割パターンに対応する英語配列(図6の例では英語配列B2～B4)を生成してもよい。

[0112] 続いて、決定部14が、分割部13により生成されたローマ字配列A及び英語配列Bに基づいて、カタカナ単語w1と英単語w2とが翻字関係にあるか否かを決定する処理を実行する(ステップS4、決定ステップ)。ここで、分割部13によって複数の分割パターンに対応する複数の英語配列(図6の例では英語配列B1～B4)が生成される場合には、ステップS4の処理は、ローマ字配列Aと英語配列B1～B4との各組み合わせについて実行される。ステップS4の処理の詳細については、図9を用いて後述する。

[0113] 続いて、決定部14によってカタカナ単語w1と英単語w2とが翻字関係にあると判定された場合(ステップS5: YES)には、出力部16がカタカナ単語w1及び英単語w2を翻字対として出力する(ステップS6、出力ステップ)。一方、決定部14によってカタカナ単語w1と英単語w2とが翻字関係にあると判定されなかった場合(ステップS5: NO)には、カタカナ単語w1及び英単語w2の組は翻字対として出力されない。

[0114] なお、ステップS1において、単語取得部11がカタカナ単語w1及び英単語w2の一方又は両方を複数取得している場合には、ステップS2～S5の処理をカタカナ単語w1及び英単語w2のすべての組み合わせについて実行してもよい。これにより、単語取得部11によって取得された複数のカタカナ単語w1及び英単語w2のうちから、翻字対を自動的に効率良く検出することができる。

[0115] 次に、図9を用いて、図8のステップS4の処理の詳細について説明する

。ここでは、上述した決定部による処理の第１～第４の例で述べた処理を組み合わせた処理について説明する。ただし、図９に示すフローチャートは、決定部１４が採用し得る処理手順の一例に過ぎない。なお、ここでは説明をわかり易くするために、上述した「決定部による処理の第３の例」と同一の例を用いて説明する。すなわち、ローマ字配列Ａ及び英語配列Ｂが、以下に示すように図６に示したローマ字配列Ａ及び英語配列Ｂ２と同一である場合を例に挙げて、決定部１４の処理を説明する。なお、ルール情報記憶部１５に記憶された子音ルール及び除外ルールは、上述した「決定部による処理の第３の例」と同様であるものとする。

A = { [“B” “A”] , [“S” “U”] , [“T” “A O”] , [“R” “U”] }

B = { [“B” “A”] , [“T H” “”] , [“T” “O”] , [“W” “E”] [“L” “”] }

[0116] まず、決定部１４は、ローマ字配列Ａの各要素の子音要素と英語配列Ｂの各要素の子音要素との比較に先立って、内部で保持するカウンタを初期化する（ステップＳ１０１）。ここで、カウンタＭは比較対象のローマ字配列Ａの要素番号を示し、カウンタＮは比較対象の英語配列Ｂの要素番号を示し、スコアＳは子音要素同士の間での比較において対応関係にあると判定された要素数を示す。ステップＳ１０１において、カウンタＭ、Ｎ及びスコアＳは、「 $M = 1$, $N = 1$, $S = 0$ 」に初期化される。

[0117] 続いて、決定部１４は、ルール情報記憶部１５に記憶された子音ルールを参照することにより、ローマ字配列Ａの要素Ａ[１]の子音要素“Ｂ”と英語配列Ｂの要素Ｂ[１]の子音要素“Ｂ”とが対応関係にあるか否かを判定する（ステップＳ１０２）。ここで、決定部１４は、要素Ａ[１]の子音要素“Ｂ”と要素Ｂ[１]の子音要素“Ｂ”とは対応関係にあると判定する（ステップＳ１０２：ＹＥＳ）。そして、決定部１４は、カウンタＭ、Ｎ及びスコアＳに１を加算する（ステップＳ１０３）。これにより、カウンタＭ、Ｎ及びスコアＳは、「 $M = 2$, $N = 2$, $S = 1$ 」に更新される。

[0118] 続いて、要素A [2] 及び要素B [2] が両方とも存在するため（ステップS 1 0 4 : Y E S）、決定部1 4は、当該比較対象の要素A [2] 及び要素B [2] の子音要素同士が対応関係にあるか否かを判定する（ステップS 1 0 2）。ここで、決定部1 4は、要素A [2] の子音要素“S”と要素B [2] の子音要素“TH”とは対応関係にあると判定する（ステップS 1 0 2 : Y E S）。そして、決定部1 4は、カウンタM, N及びスコアSに1を加算する（ステップS 1 0 3）。これにより、カウンタM, N及びスコアSは、「M=3, N=3, S=2」に更新される。

[0119] 続いて、要素A [3] 及び要素B [3] が両方とも存在するため（ステップS 1 0 4 : Y E S）、決定部1 4は、当該比較対象の要素A [3] 及び要素B [3] の子音要素同士が対応関係にあるか否かを判定する（ステップS 1 0 2）。ここで、決定部1 4は、要素A [3] の子音要素“T”と要素B [3] の子音要素“T”とは対応関係にあると判定する（ステップS 1 0 2 : Y E S）。そして、決定部1 4は、カウンタM, N及びスコアSに1を加算する（ステップS 1 0 3）。これにより、カウンタM, N及びスコアSは、「M=4, N=4, S=3」に更新される。

[0120] 続いて、要素A [4] 及び要素B [4] が両方とも存在するため（ステップS 1 0 4 : Y E S）、決定部1 4は、当該比較対象の要素A [4] 及び要素B [4] の子音要素同士が対応関係にあるか否かを判定する（ステップS 1 0 2）。ここで、決定部1 4は、要素A [4] の子音要素“R”と要素B [4] の子音要素“W”とは対応関係にないと判定する（ステップS 1 0 2 : N O）。このため、決定部1 4は、ルール情報記憶部1 5に記憶された除外ルールを参照することにより、要素B [4] の子音要素を比較対象から除外可能か否かを判定する（ステップS 1 0 5）。

[0121] 要素B [4] の子音要素“W”は除外ルールに示される条件に合致するため、決定部1 4は、要素B [4] の子音要素が除外可能であると判定する（ステップS 1 0 5 : Y E S）。そして、決定部1 4は、要素B [4] を比較対象から除外するためにカウンタNに1を加算し、「N=5」に更新する（

ステップS 1 0 6)。そして、決定部 1 4 は、次の要素 B [5] が存在するか否かを判定する（ステップS 1 0 7）。

[0122] ここで、要素 B [5] は存在するため（ステップS 1 0 7 : Y E S）、決定部 1 4 は、ステップS 1 0 2に戻って、要素 A [4] の子音要素 “R” と要素 B [5] の子音要素 “L” とが対応関係にあるか否かを判定する。ここで、決定部 1 4 は、要素 A [4] の子音要素 “R” と要素 B [5] の子音要素 “L” とは対応関係にあると判定する（ステップS 1 0 2 : Y E S）。そして、決定部 1 4 は、カウンタ M, N 及びスコア S に 1 を加算する（ステップS 1 0 3）。これにより、カウンタ M, N 及びスコア S は、「M = 5, N = 6, S = 4」に更新される。

[0123] 続いて、要素 A [5] 及び要素 B [6] はいずれも存在しないため（ステップS 1 0 4 : N O、ステップS 1 0 8 : N O）、決定部 1 4 は、この時点で、ローマ字配列 A の各要素の子音要素と英語配列 B の各要素の子音要素とが対応関係にあると判定する。その上で、決定部 1 4 は、ローマ字配列 A 及び英語配列 B について母音要素の対応関係を判定するか否かを決定する（ステップS 1 1 0）。

[0124] 上述した通り、ステップS 1 1 0において、決定部 1 4 は、例えばスコア S が予め定めた閾値以下である場合に母音要素の対応関係を判定すると決定することができる。ここで、仮に閾値が「3」として定められている場合、スコア S (= 4) は当該閾値より大きいため、決定部 1 4 は、母音要素の対応関係を判定しないと決定する（ステップS 1 1 0 : N O）。この場合、決定部 1 4 は、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にあると決定して処理を終了する（ステップS 1 1 4）。

[0125] 以下、図 9 に示すフローチャートにおいて、上記具体例において説明できなかった部分について補足する。

[0126] ステップS 1 0 5において、要素 B [N] の子音要素が除外可能であると判定されなかった場合（ステップS 1 0 5 : N O）には、決定部 1 4 は、子音要素同士の判定に失敗した（すなわち、ローマ字配列 A の各要素の子音要

素と英語配列 B の各要素の子音要素とが対応関係にない) と判定する。この場合、決定部 14 は、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にないとして処理を終了する (ステップ S 113)。

[0127] ステップ S 107 において、要素 B [N] が存在しない場合 (ステップ S 107 : NO) とは、ローマ字配列 A の要素 A [M] と比較するための英語配列 B の要素が残っていない場合を意味する。この場合、少なくとも要素 A [M] の子音要素に対応する子音要素が英語配列 B には存在しないこととなるため、決定部 14 は、子音要素同士の判定に失敗したと判定する。この場合、決定部 14 は、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にないとして処理を終了する (ステップ S 113)。

[0128] ステップ S 108 において、要素 A [M] 及び要素 B [N] のいずれか一方が存在する場合 (ステップ S 108 : YES)、決定部 14 は、ルール情報記憶部 15 に記憶された除外ルールを参照することにより、残存する要素の子音要素が除外可能か否かを判定する (ステップ S 109)。例えば、上述した「決定部による処理の第 2 の例」の場合 (図 5 の例の場合)、英語配列 B の最後の要素 B [4] (= [“S” “”]) のみが残ることになる。

[0129] このような場合において、残存する要素の子音要素が除外可能であると判定された場合 (ステップ S 109 : YES)、決定部 14 は、ローマ字配列 A の各要素の子音要素と英語配列 B の各要素の子音要素とが対応関係にあると判定する。その上で、決定部 14 は、ローマ字配列 A 及び英語配列 B について母音要素の対応関係を判定するか否かを決定する (ステップ S 110)。

[0130] 一方、残存する要素の子音要素が除外可能であると判定されなかった場合 (ステップ S 109 : NO)、残存する要素の子音要素に関する比較を実行することができないため、決定部 14 は、子音要素同士の判定に失敗したと判定する。すなわち、決定部 14 は、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にないとして処理を終了する (ステップ S 113)。

[0131] ステップ S 110 において、母音要素の対応関係を判定すると決定された

場合（ステップS 1 1 0：YES）、決定部 1 4 は、上述の「決定部による処理の第 4 の例」で述べたような処理を実行する。すなわち、決定部 1 4 は、ルール情報記憶部 1 5 に記憶された禁止ルールを参照することにより、ローマ字配列 A の各要素の母音要素と英語配列 B の各要素の母音要素との各組み合わせが禁止ルールに該当するか否かを判定する（ステップS 1 1 2）。ここで、いずれかの組み合わせについて禁止ルールに該当すると判定された場合（ステップS 1 1 2：YES）、決定部 1 4 は、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にないと決定する（ステップS 1 1 3）。一方、いずれの組み合わせについても禁止ルールに該当すると判定されなかった場合（ステップS 1 1 2：NO）、決定部 1 4 は、カタカナ単語 w 1 と英単語 w 2 とが翻字関係にあると決定する（ステップS 1 1 4）。

[0132] 次に、図 1 0 を用いて、コンピュータ 1 0 0 を翻字処理装置 1 0 として機能させるための翻字処理プログラム P 1 について説明する。

[0133] 翻字処理プログラム P 1 は、単語取得モジュール P 1 1、文字列取得モジュール P 1 2、分割モジュール P 1 3、決定モジュール P 1 4、及び出力モジュール P 1 5 を備える。単語取得モジュール P 1 1、文字列取得モジュール P 1 2、分割モジュール P 1 3、決定モジュール P 1 4、及び出力モジュール P 1 5 を実行することにより実現される機能はそれぞれ、上記の単語取得部 1 1、文字列取得部 1 2、分割部 1 3、決定部 1 4、及び出力部 1 6 の機能と同様である。

[0134] 翻字処理プログラム P 1 は、例えば、CD-ROM や DVD-ROM、半導体メモリ等の有形の記録媒体に固定的に記録された上で提供されてもよい。或いは、翻字処理プログラム P 1 は、搬送波に重畳されたデータ信号として通信ネットワークを介して提供されてもよい。

[0135] 以上説明したように、本発明の一側面に係る翻字処理装置は、所定の文字体系を有する第 1 の言語で記述された第 1 の単語をアルファベット表記で表した第 1 のアルファベット文字列と、第 1 の言語とは異なる文字体系を有する第 2 の言語で記述された第 2 の単語をアルファベット表記で表した第 2 の

アルファベット文字列と、を取得する文字列取得部と、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて第1の単語と第2の単語とが翻字関係にあるか否かを決定する決定部と、決定部によって翻字関係にあると決定された第1の単語及び第2の単語を翻字対として出力する出力部と、を備える。

[0136] 本発明の一側面に係る翻字処理方法は、少なくとも一つのプロセッサにより実行される翻字処理方法であって、所定の文字体系を有する第1の言語で記述された第1の単語をアルファベット表記で表した第1のアルファベット文字列と、第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得ステップと、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて第1の単語と第2の単語とが翻字関係にあるか否かを決定する決定ステップと、決定ステップにおいて翻字関係にあると決定された第1の単語及び第2の単語を翻字対として出力する出力ステップと、を含む。

[0137] 本発明の一側面に係る翻字処理プログラムは、所定の文字体系を有する第1の言語で記述された第1の単語をアルファベット表記で表した第1のアルファベット文字列と、第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得ステップと、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて第1の単語と第2の単語とが翻字関係にあるか否かを決定する決定ステップと、決定ステップにおいて翻字関係にあると決定された第1の単語及び第2の単語を翻字対として出力する出力ステップと、をコンピュータに実行させる。

- [0138] このような側面においては、第1のアルファベット文字列に含まれる第1の子音要素と第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定結果に基づいて、第1の言語で記述された第1の単語と第2の言語で記述された第2の単語とが翻字関係にあるか否かが決定される。このように、互いに異なる文字体系を有する言語で記述された単語同士が翻字関係にあるか否かを、それぞれの単語をアルファベット表記で表したアルファベット文字列に含まれる子音要素間の対応関係に着目して決定することにより、翻字対を効率良く検出することができる。
- [0139] 他の側面に係る翻字処理装置では、決定部は、第1のアルファベット文字列に含まれる第1の子音要素及び第2のアルファベット文字列に含まれる第2の子音要素のうち、予め定めた条件に合致する子音要素を除外し、残される子音要素について上記判定を実行してもよい。
- [0140] 第1の単語と第2の単語とが翻字関係にある場合において、一方の単語のアルファベット文字列に現れる子音要素に対応する子音要素が、他方の単語のアルファベット文字列には明示的に現れない場合がある。そこで、決定部は、このような子音要素を予め定めた条件に合致するか否かの判定によって抽出し、抽出された子音要素を除外することによって、第1の単語と第2の単語とが翻字関係にあるか否かをより精度良く決定することが可能となる。
- [0141] 他の側面に係る翻字処理装置では、第1のアルファベット文字列を母音字から子音字に切り替わる境界で分割し、分割された個々の要素を第1のアルファベット文字列における出現順に配列した第1の配列を生成するとともに、第2のアルファベット文字列を母音字から子音字に切り替わる境界で分割し、分割された個々の要素を第2のアルファベット文字列における出現順に配列した第2の配列を生成する分割部を更に備え、決定部は、第1の配列の各要素と第2の配列の各要素とを先頭の要素から順に要素ごとに比較し、第1の配列の各要素に含まれる第1の子音要素と第2の配列の各要素に含まれる第2の子音要素とが予め定められた対応関係にあると判定された場合に、第1の単語と第2の単語とが翻字関係にあると決定してもよい。

[0142] 第1のアルファベット文字列に含まれる各子音要素と第2のアルファベット文字列に含まれる各子音要素とが対応関係にある場合には、第1の単語と第2の単語とが翻字関係にある可能性が高い。従って、上述した構成によれば、第1の単語と第2の単語とが翻字関係にあるか否かを、子音要素同士の対応関係に基づいて容易に切り分けることができ、翻字対を効率良く検出することができる。

[0143] 他の側面に係る翻字処理装置では、分割部は、分割された個々の要素のうちに複数の子音字からなる子音要素が含まれている要素がある場合に、当該要素を二以上の要素に更に分割する場合と分割しない場合とを含む複数の分割パターンの各々に対応する第1の配列及び第2の配列を生成し、決定部は、複数の分割パターンの各々について、第1の配列の各要素と第2の配列の各要素とを先頭の要素から順に要素ごとに比較し、少なくとも一つの分割パターンについて第1の配列の各要素に含まれる第1の子音要素と第2の配列の各要素に含まれる第2の子音要素とが予め定められた対応関係にあると判定された場合に、第1の単語と第2の単語とが翻字関係にあると決定してもよい。

[0144] 第1のアルファベット文字列及び第2のアルファベット文字列の一方の文字列に含まれる複数の子音字からなる一つの子音要素は、他方の文字列において母音要素によって分断される二以上の子音要素に対応する場合がある。従って、複数の子音字からなる子音要素を更に分割するパターンと分割しないパターンとを含む複数の分割パターンの各々についての判定を実行することにより、第1の単語と第2の単語とが翻字関係にあるか否かを精度良く検出することが可能となる。

[0145] 他の側面に係る翻字処理装置では、決定部は、第1のアルファベット文字列に含まれる第1の母音要素と第2のアルファベット文字列に含まれる第2の母音要素との組み合わせに基づく判定を更に実行し、当該判定の結果にも基づいて第1の単語と第2の単語とが翻字関係にあるか否かを決定してもよい。

- [0146] 子音要素同士の比較に加えて、更に母音要素同士の比較にも基づくことで、第1の単語と第2の単語とが翻字関係にあるか否かを精度良く決定することが可能となる。
- [0147] 他の側面に係る翻字処理装置では、一のウェブページに含まれるカタカナ表記の単語を第1の単語として取得し、一のウェブページに含まれるアルファベット表記の単語を第2の単語として取得し、取得された第1の単語及び第2の単語を文字列取得部に受け渡す単語取得部を更に備えてもよい。
- [0148] 同一のウェブページに含まれるテキスト情報には、同一の概念についてカタカナ表記及びアルファベット表記のそれぞれで表した単語が含まれている可能性が高い。言い換えれば、同一のウェブページに含まれるテキスト情報には、互いに翻字関係にあるカタカナ表記の単語及びアルファベット表記の単語が含まれている可能性が高い。従って、同一のウェブページに含まれるカタカナ表記の単語及びアルファベット表記の単語を第1の単語及び第2の単語として取得することにより、翻字対を効率良く検出することが期待できる。
- [0149] また、他の側面に係る翻字処理装置では、一のウェブページは、商品又はサービスをインターネット上で販売するウェブサイトにおいて商品又はサービスを提示するウェブページであり、単語取得部は、一のウェブページにおいて販売される商品又はサービスの名称の記載箇所を特定し、当該記載箇所に含まれるカタカナ表記で表された単語を第1の単語として取得し、当該記載箇所に含まれるアルファベット表記で表された単語を第2の単語として取得してもよい。
- [0150] 商品又はサービスをインターネット上で販売するウェブサイトにおいて商品又はサービスを提示するウェブページには、商品又はサービスの名称の記載箇所が設けられることがある。ここで、商品又はサービスの名称が、カタカナ表記で表される場合（例えば、商品又はサービスの名称が、外来語で表される物の名称、外国の地名、外国人の氏名等に基づくものである場合）には、商品又はサービスの名称のカタカナ表記及びアルファベット表記の両方

が上記記載箇所に含まれている可能性が高い。従って、上記記載箇所から、翻字関係にある可能性の高いカタカナ表記で表された単語及びアルファベット表記で表された単語を第１の単語及び第２の単語として取得することで、翻字対をより効率良く検出することが期待できる。

[0151] 以上、本発明をその実施形態に基づいて詳細に説明した。しかし、本発明は上記実施形態に限定されるものではない。本発明は、その要旨を逸脱しない範囲で様々な変形が可能である。

[0152] 上記実施形態では、分割部１３の処理の例として第１及び第２の例を説明したが、分割部１３はこれらの例で示した処理のうちの一部を実行するように構成されてもよい。また、上記実施形態では、決定部１４の処理の例として第１～第４の例を説明したが、決定部１４はこれらの例で示した処理のうちの一部を実行するように構成されてもよい。

[0153] 例えば、決定部１４は、第２の例の処理（子音要素の除外に関する処理）を実行する構成を備えなくともよい。この場合、図９に示したフローは、ステップＳ１０５～Ｓ１０７とステップＳ１０９とを省略したフローに書き換えられる。この場合において、ステップＳ１０８の判定がＹＥＳであった場合にはステップＳ１１３に進む。

[0154] また、分割部１３は、第２の例の処理（複数の分割パターンに対応する配列を生成する処理）を実行する構成を備えなくともよい。これに対応して、決定部１４は、第３の例の処理（複数の分割パターンのそれぞれに対応する比較を実行する処理）を実行する構成を備えなくともよい。

[0155] また、決定部１４は、第４の例の処理（母音要素の組み合わせに基づく判定処理）を実行する構成を備えなくともよい。この場合、図９に示したフローは、ステップＳ１１０～Ｓ１１２を省略したフローに書き換えられる。

[0156] 上記で述べたいずれの変形例においても、カタカナ単語ｗ１と英単語ｗ２とが翻字関係にあるか否かを、子音要素同士に対応関係に基づいて容易に切り分けることができ、翻字対を効率良く検出することができるという効果が奏される。

[0157] また、上述した通り、少なくとも一つのプロセッサにより実行される翻字処理方法の処理手順は図8及び図9の例に限定されない。例えば、翻字処理装置10は、これらの図におけるステップ（処理）の一部を省略してもよいし、別の順序で各ステップの処理を実行してもよい。また、これらの図に示されるステップの任意の2以上のステップが組み合わされてもよいし、ステップの一部が修正または削除されてもよい。或いは、翻字処理装置10は、これらの図に示される各ステップに加えて他のステップを実行してもよい。

[0158] 情報処理システム1内で二つの数値の大小関係を比較する際には、「以上」及び「よりも大きい」という二つの基準のどちらを用いてもよく、「以下」及び「未満」の二つの基準のうちのどちらを用いてもよい。このような基準の選択は、二つの数値の大小関係を比較する処理についての技術的意義を変更するものではない。

符号の説明

[0159] 1…情報処理システム、10…翻字処理装置、11…単語取得部、12…文字列取得部、13…分割部、14…決定部、15…ルール情報記憶部、16…出力部、20…情報処理装置、P1…翻字処理プログラム、P11…単語取得モジュール、P12…文字列取得モジュール、P13…分割モジュール、P14…決定モジュール、P15…出力モジュール。

請求の範囲

[請求項1] 所定の文字体系を有する第1の言語で記述された第1の単語をアルファベット表記で表した第1のアルファベット文字列と、前記第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得部と、

前記第1のアルファベット文字列に含まれる第1の子音要素と前記第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて前記第1の単語と前記第2の単語とが翻字関係にあるか否かを決定する決定部と、

前記決定部によって翻字関係にあると決定された前記第1の単語及び前記第2の単語を翻字対として出力する出力部と、
を備える翻字処理装置。

[請求項2] 前記決定部は、前記第1のアルファベット文字列に含まれる前記第1の子音要素及び前記第2のアルファベット文字列に含まれる前記第2の子音要素のうち、予め定めた条件に合致する子音要素を除外し、残される子音要素について前記判定を実行する、
請求項1に記載の翻字処理装置。

[請求項3] 前記第1のアルファベット文字列を母音字から子音字に切り替わる境界で分割し、分割された個々の要素を前記第1のアルファベット文字列における出現順に配列した第1の配列を生成するとともに、前記第2のアルファベット文字列を母音字から子音字に切り替わる境界で分割し、分割された個々の要素を前記第2のアルファベット文字列における出現順に配列した第2の配列を生成する分割部を更に備え、

前記決定部は、前記第1の配列の各要素と前記第2の配列の各要素とを先頭の要素から順に要素ごとに比較し、前記第1の配列の各要素に含まれる前記第1の子音要素と前記第2の配列の各要素に含まれる

前記第 2 の子音要素とが前記予め定められた対応関係にあると判定された場合に、前記第 1 の単語と前記第 2 の単語とが翻字関係にあると決定する、

請求項 1 又は 2 に記載の翻字処理装置。

[請求項 4]

前記分割部は、前記分割された個々の要素のうちに複数の子音字からなる子音要素が含まれている要素がある場合に、当該要素を二以上の要素に更に分割する場合と分割しない場合とを含む複数の分割パターンの各々に対応する前記第 1 の配列及び前記第 2 の配列を生成し、

前記決定部は、前記複数の分割パターンの各々について、前記第 1 の配列の各要素と前記第 2 の配列の各要素とを先頭の要素から順に要素ごとに比較し、少なくとも一つの前記分割パターンについて前記第 1 の配列の各要素に含まれる前記第 1 の子音要素と前記第 2 の配列の各要素に含まれる前記第 2 の子音要素とが前記予め定められた対応関係にあると判定された場合に、前記第 1 の単語と前記第 2 の単語とが翻字関係にあると決定する、

請求項 3 に記載の翻字処理装置。

[請求項 5]

前記決定部は、前記第 1 のアルファベット文字列に含まれる第 1 の母音要素と前記第 2 のアルファベット文字列に含まれる第 2 の母音要素との組み合わせに基づく判定を更に実行し、当該判定の結果にも更に基づいて前記第 1 の単語と前記第 2 の単語とが翻字関係にあるか否かを決定する、

請求項 1 ～ 4 のいずれか一項に記載の翻字処理装置。

[請求項 6]

一のウェブページに含まれるカタカナ表記の単語を前記第 1 の単語として取得し、前記一のウェブページに含まれるアルファベット表記の単語を前記第 2 の単語として取得し、取得された前記第 1 の単語及び前記第 2 の単語を前記文字列取得部に受け渡す単語取得部を更に備える、

請求項 1 ～ 5 のいずれか一項に記載の翻字処理装置。

[請求項7] 前記一のウェブページは、商品又はサービスをインターネット上で販売するウェブサイトにおいて商品又はサービスを提示するウェブページであり、

前記単語取得部は、前記一のウェブページにおいて販売される商品又はサービスの名称の記載箇所を特定し、当該記載箇所に含まれるカタカナ表記で表された単語を前記第1の単語として取得し、当該記載箇所に含まれるアルファベット表記で表された単語を前記第2の単語として取得する、

請求項6に記載の翻字処理装置。

[請求項8] 請求項1～7のいずれか一項に記載の翻字処理装置により出力された前記翻字対を参照することにより予め定められた処理を実行する情報処理装置。

[請求項9] 少なくとも一つのプロセッサにより実行される翻字処理方法であって、

所定の文字体系を有する第1の言語で記述された第1の単語をアルファベット表記で表した第1のアルファベット文字列と、前記第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得ステップと、

前記第1のアルファベット文字列に含まれる第1の子音要素と前記第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて前記第1の単語と前記第2の単語とが翻字関係にあるか否かを決定する決定ステップと、

前記決定ステップにおいて翻字関係にあると決定された前記第1の単語及び前記第2の単語を翻字対として出力する出力ステップと、を含む翻字処理方法。

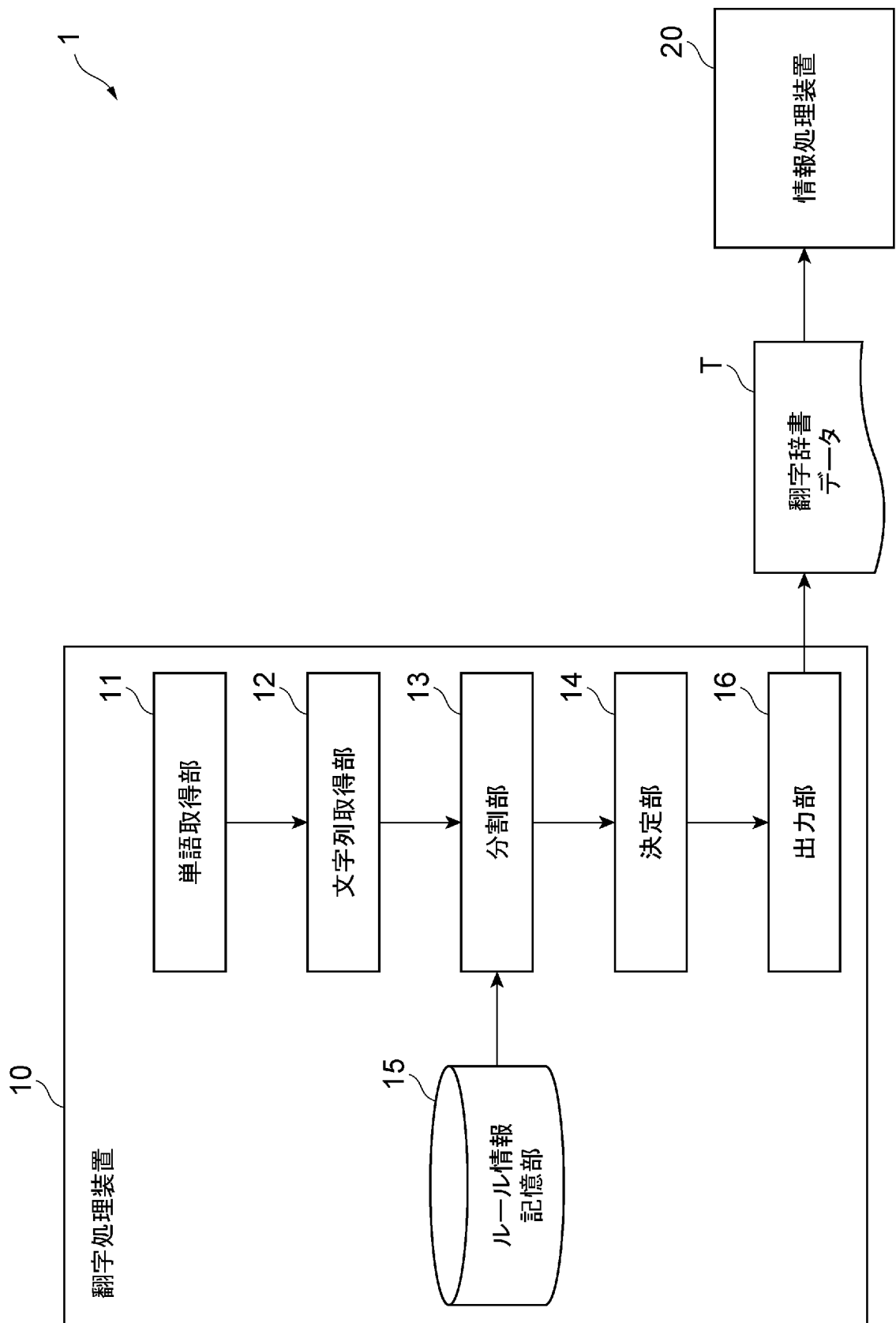
[請求項10] 所定の文字体系を有する第1の言語で記述された第1の単語をアル

ファベット表記で表した第1のアルファベット文字列と、前記第1の言語とは異なる文字体系を有する第2の言語で記述された第2の単語をアルファベット表記で表した第2のアルファベット文字列と、を取得する文字列取得ステップと、

前記第1のアルファベット文字列に含まれる第1の子音要素と前記第2のアルファベット文字列に含まれる第2の子音要素とが予め定められた対応関係にあるか否かの判定を実行し、当該判定の結果に基づいて前記第1の単語と前記第2の単語とが翻字関係にあるか否かを決定する決定ステップと、

前記決定ステップにおいて翻字関係にあると決定された前記第1の単語及び前記第2の単語を翻字対として出力する出力ステップと、をコンピュータに実行させる翻字処理プログラム。

[図1]

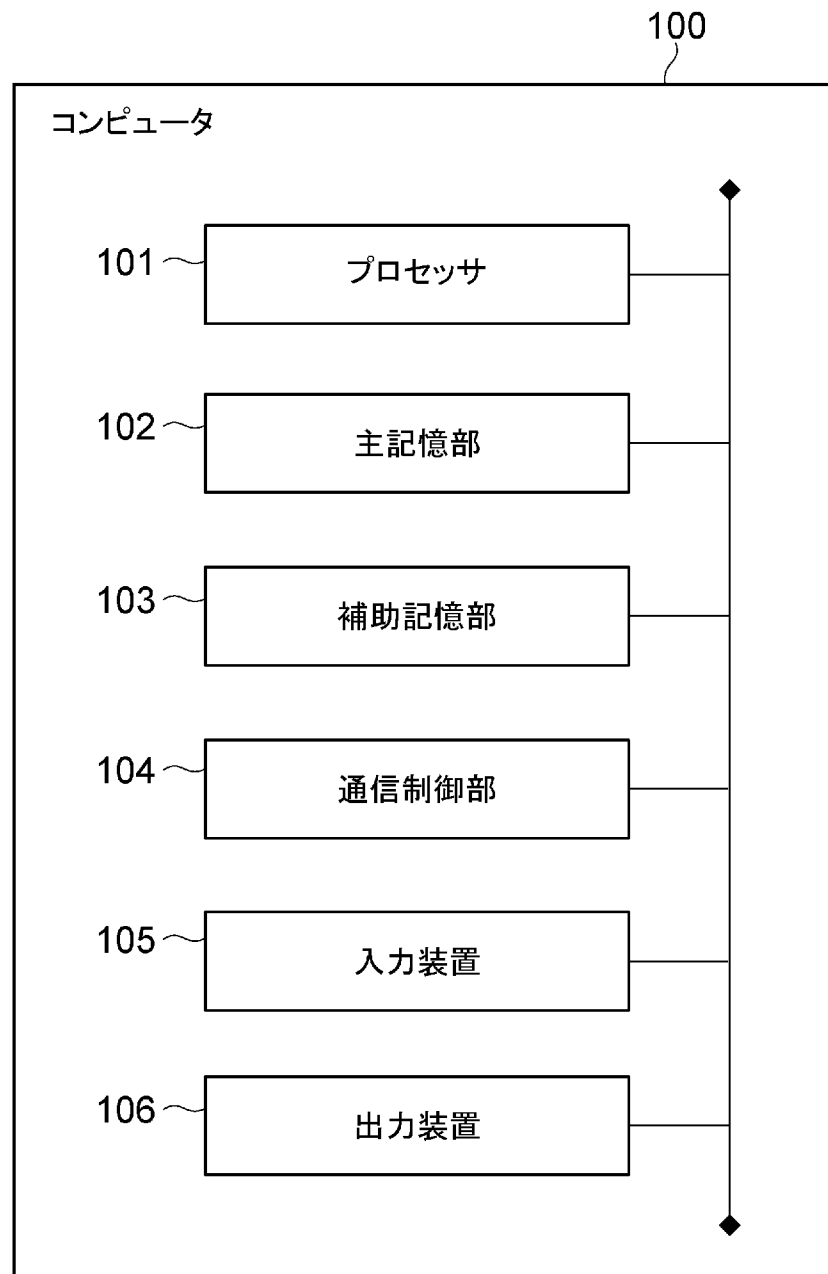


[図2]

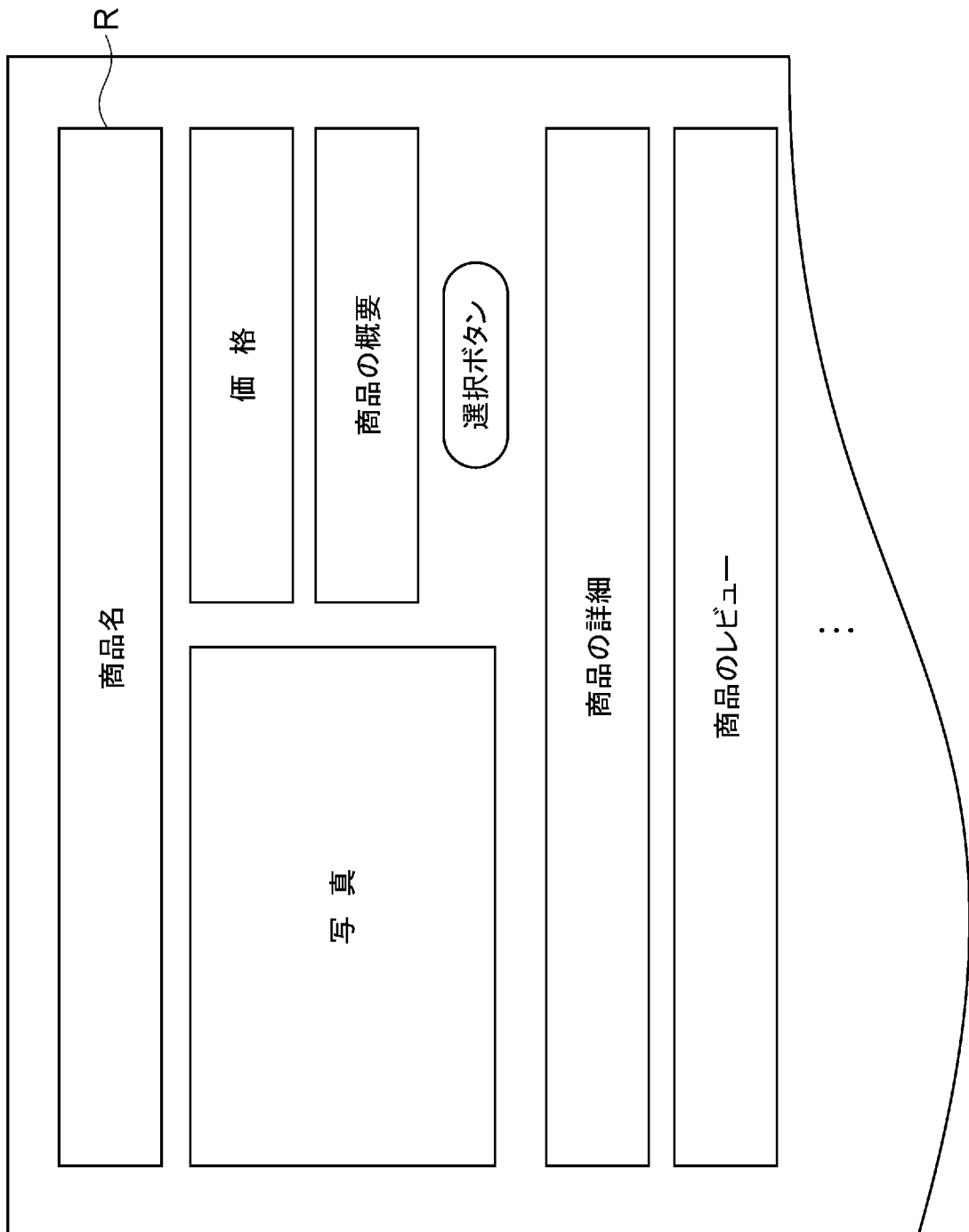
カタカナ単語(ローマ字表記)	英単語
ボジョレー (BOJOREE)	BEAUJOLAIS
プリムール (PURIMUURU)	PRIMEUR
バスタオル (BASUTAORU)	BATHTOWEL
⋮	⋮

T

[図3]



[図4]



[図5]

(a)

 $w1 = \text{「ボジョレー」}$  $str1 = \text{“BOJOREE”}$  $A = \{ [\text{“B” “O”}], [\text{“J” “O”}], [\text{“R” “EE”}] \}$

(b)

 $w2 = \text{「BEAUJOLAIS」}$  $str2 = \text{“BEAUJOLAIS”}$  $B = \{ [\text{“B” “EAU”}], [\text{“J” “O”}], [\text{“L” “AI”}], [\text{“S” “”}] \}$

[図6]

(a)

w1 = 「バスタオル」



str1 = "BASUTAORU"



$$A = \{ ["B" "A"], ["S" "U"], ["T" "AO"], ["R", "U"] \}$$

(b)

w2 = 「BATHTOWEL」



str2 = "BATHTOWEL"



$$B1 = \{ ["B" "A"], ["THT" "O"], ["W" "E"], ["L" ""] \}$$

$$B2 = \{ ["B" "A"], ["TH" " "], ["T" "O"], ["W" "E"], ["L" ""] \}$$

$$B3 = \{ ["B" "A"], ["T" " "], ["HT" "O"], ["W" "E"], ["L" ""] \}$$

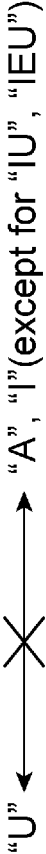
$$B4 = \{ ["B" "A"], ["T" " "], ["H" " "], ["T" "O"], ["W" "E"], ["L" ""] \}$$

[図7]

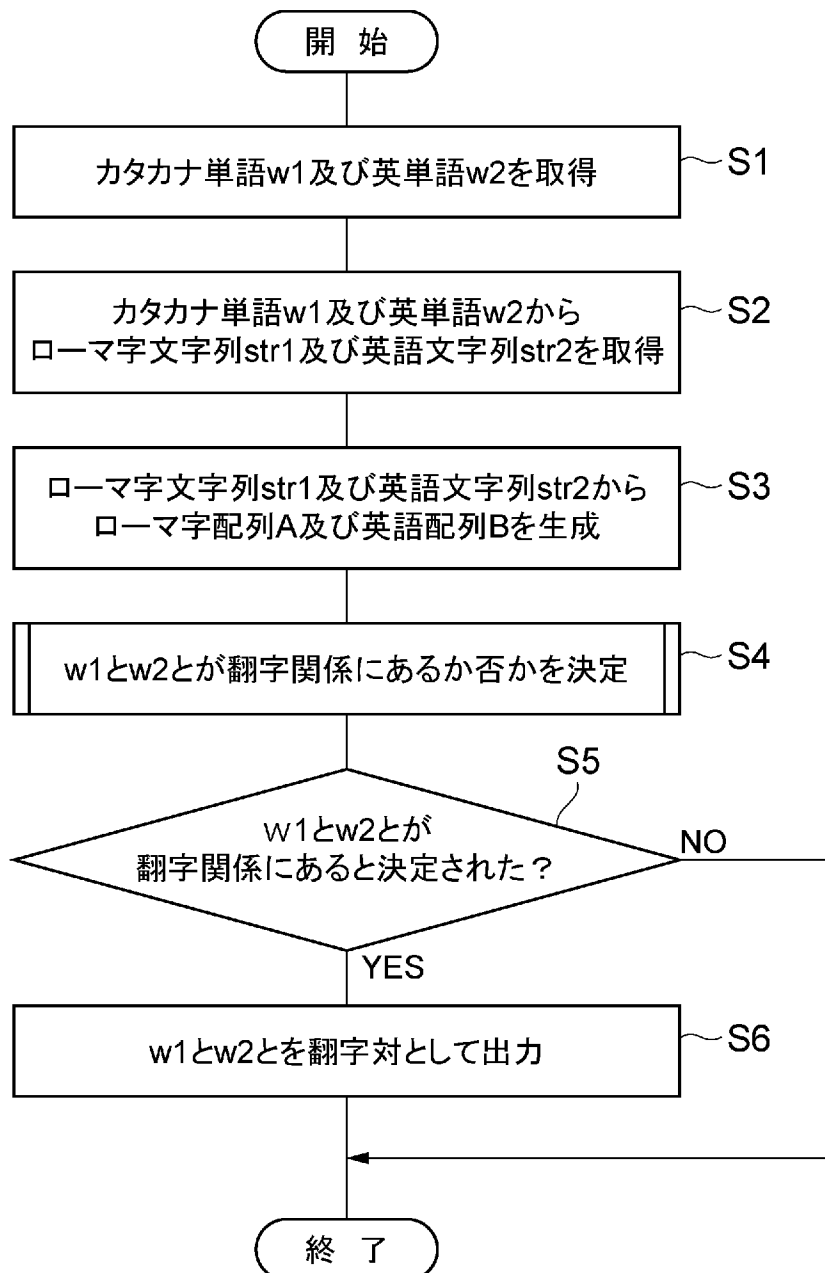
(a)

日本語(ローマ字)の子音要素	英語の子音要素
"B" "BY"	"V" "VV" "W" "B" "BB" "BH" "BJ" "BY"
"CH"	"T" "TT" "TCH" "CH" "CCH" "C" "CC"
"D" "DY"	"D" "DD" "DH" "T" "DY" "DJ"
"R" "RY"	"L" "LL" "R" "RR" "RY" "LY" "GL" "TL", "LH" "RH" "SL" "DL" "GR"
"T"	"CH" "TS" "TTS" "TZ" "T" "TT" "TH", "DT" "CHT" "D"
⋮	⋮

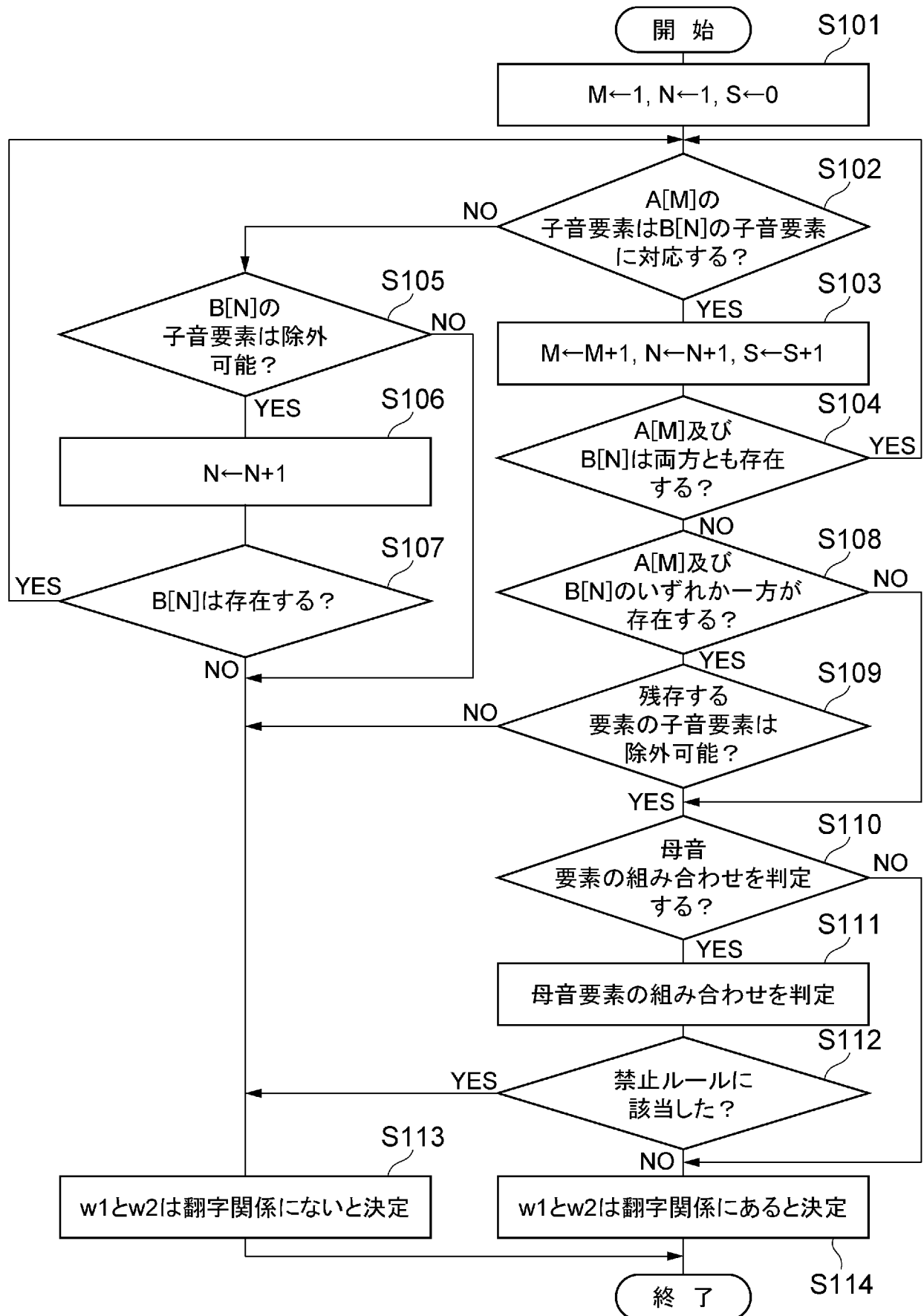
(b)



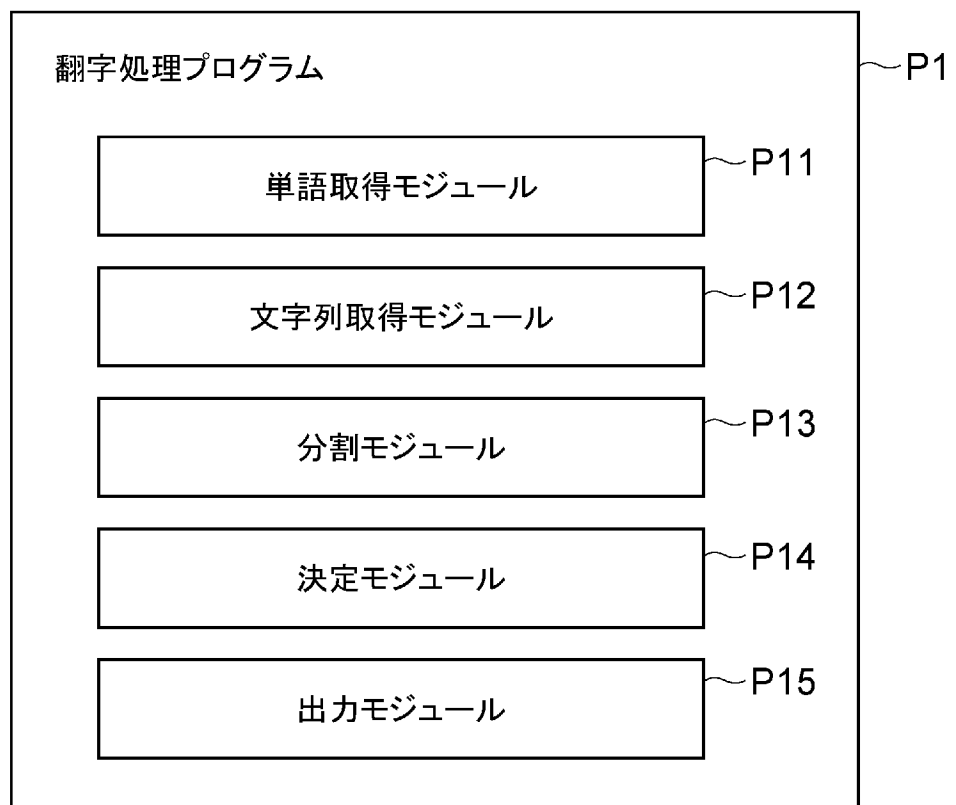
[図8]



[図9]



[図10]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/068828

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/27(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)

G06F17/27

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

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.
Y A	JP 10-143514 A (Nippon Telegraph and Telephone Corp.), 29 May 1998 (29.05.1998), paragraphs [0013], [0019] to [0024]; fig. 1, 5, 6 (Family: none)	1, 2, 5, 6, 9, 10 3, 4, 7, 8
Y A	JP 2003-263429 A (Nippon Hoso Kyokai), 19 September 2003 (19.09.2003), paragraphs [0014] to [0022]; fig. 2 (Family: none)	1, 2, 5, 6, 9, 10 3, 4, 7, 8
Y A	Yoshihiro MATSUO, Satoshi SHIRAI, "Using Pronunciation to Automatically Extract Bilingual Word Pairs", IPSJ SIG Notes, 19 November 1996 (19.11.1996), vol.96, no.114, pages 101 to 106	2, 5, 6 3, 4, 7

☒ 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
30 July 2015 (30.07.15)

Date of mailing of the international search report
11 August 2015 (11.08.15)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/068828

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	WO 2012/121063 A1 (Rakuten, Inc.), 13 September 2012 (13.09.2012), paragraph [0029] & US 2013/0246042 A1 & EP 2672395 A1 & CA 2813218 A1 & KR 20130038959 A & CN 103140849 A & TW 201250498 A	8 1-7, 9, 10

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F17/27(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（I P C））

Int.Cl. G06F17/27

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	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 5 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 5 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 5 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
Y A	JP 10-143514 A（日本電信電話株式会社）1998.05.29, 段落 [0013]、[0019] - [0024] 及び図1, 5, 6 (ファミリーなし)	1, 2, 5, 6, 9, 10 3, 4, 7, 8
Y A	JP 2003-263429 A（日本放送協会）2003.09.19, 段落 [0014] - [0022] 及び図2（ファミリーなし）	1, 2, 5, 6, 9, 10 3, 4, 7, 8

☒ 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

3 0 . 0 7 . 2 0 1 5

国際調査報告の発送日

1 1 . 0 8 . 2 0 1 5

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

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）

川▲崎▼ 博章

電話番号 03-3581-1101 内線 3599

5 M

5 2 8 4

C (続き) . 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
Y A	松尾 義博、白井 諭, 発音情報を用いた訳語対の自動抽出, 情報処理学会研究報告, 1996. 11. 19, V o l . 9 6、N o . 1 1 4, p . 1 0 1－1 0 6	2, 5, 6 3, 4, 7
X A	WO 2012/121063 A1 (楽天株式会社) 2012. 09. 13, 段落 [0029] & US 2013/0246042 A1 & EP 2672395 A1 & CA 2813218 A1 & KR 20130038959 A & CN 103140849 A & TW 201250498 A	8 1－7, 9, 10